DOCUMENT RESUME

ED 375 157 TM 022 107

TITLE	Committee Meeting of Assembly Education Committee "To Receive Testimony from the Commissioner of Education, Mary Lee Fitzgerald, Department Staff, and Others Concerning the Department's Skills Testing Program, Including the Early Warning Test and High School Proficiency Test, Pursuant to Assembly Resolution No. 113."
INSTITUTION	New Jersey State Office of Legislative Services, Trenton. Assembly Education Committee.
PUB DATE	19 Apr 93
NOTE	209p.; For a related document, see TM 022 106. Some pages may not reproduce well.
PUB TYPE	Legal/Legislative/Regulatory Materials (090)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	Accountability; Achievement Tests; *Basic Skills; Cost Effectiveness; *Educational Change; Elementary Secondary Education; Grade 8; Legislation; Remedial Instruction; Standards; *State Programs; Student Motivation; Student Placement; Testing Problems; *Testing Programs; Test Use
IDENTIFIERS	*New Jersey; Reform Efforts; State Legislatures

ABSTRACT

The Assembly Education Committee of the New Jersey Office of Legislative Services held a hearing pursuant to Assembly Resolution 113, a proposal directing the Committee to investigate the skills testing program developed and administered to New Jersey children by the State Department of Education. The Committee was interested in the eighth-grade early warning test designed to identify students who might have problems with the proposed High School Proficiency Test. Testifying about the testing programs were: (1) Mary Lee Fitzgerald, Commissioner, New Jersey Department of Education; (2) Leo Klagholz, Assistant Commissioner, Division of Academic Programs and Standards, New Jersey Department of Education; (3) Patricia Wang Iverson, private citizen; (4) Ellen Oppenheimer, private citizen; (5) Joseph G. Rosenstein, Ph.D., Mathematics Department, Rutgers - The State University, and Director, New Jersey Mathematics Coalition; (6) Patricia Clark Kenschaft, Ph.D. Professor of Mathematics, Montclair State College and Director, Project for Resourceful Instruction of Mathematics in the Elementary School-PRIMES; and (7) Ira Sweet, teacher, guidance counselor and school psychologist in the New Jersey Public School System. Several supplemental reports and correspondence introduced by those testifying are appended. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Tm

Committee Meeting

of

ASSEMBLY EDUCATION COMMITTEE

"To receive testimony from the Commissioner of Education, Mary Lee Fitzgerald, Department staff, and others concerning the Department's skills testing program, including the early warning test and high school proficiency test, pursuant to Assembly Resolution No. 113"

ED 375 157

LOCATION: Committee Room 8
Legislative Office Building
Trenton, New Jersey

DATE: April 19, 1993
9:55 a.m.

MEMBERS OF COMMITTEE PRESENT:

Assemblyman John A. Rocco, Chairman
Assemblyman Robert J. Martin, Vice-Chairman
Assemblyman Fredrick P. Nickles
Assemblyman David W. Wolfe
Assemblyman Joseph Charles, Jr.
Assemblyman Raul "Rudy" Garcia

ALSO PRESENT:

Senator John H. Ewing, District 16
Assemblyman Joseph V. Doria, Jr., District 31
Assemblywoman Priscilla B. Anderson, District 7

David C. Hespe
Office of Legislative Services
Aide, Assembly Education Committee

Hearing Recorded and Transcribed by
The Office of Legislative Services, Public Information Office,
Hearing Unit, State House Annex, CN 068, Trenton, New Jersey 08625

TM022107

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
SUSAN STERNBERG

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

JOHN A. ROCCO
Chairman
ROBERT J. MARTEN
Vice-Chairman
FREDRICK P. NICKLES
MARY VIRGINIA "GINNY" WEBER
DAVID W. WOLFE
Joseph Charles, Jr.
Raul "Rudy" Garcia

New Jersey State Legislature

ASSEMBLY EDUCATION COMMITTEE
LEGISLATIVE OFFICE BUILDING, CN-068
TRENTON, NEW JERSEY 08625-0068
(609) 984-6843

COMMITTEE NOTICE

TO: MEMBERS OF THE ASSEMBLY EDUCATION COMMITTEE

FROM: ASSEMBLYMAN JOHN A. ROCCO, CHAIRMAN

SUBJECT: **COMMITTEE MEETING - April 19, 1993**

The public may address comments and questions to David C. Hespe, Committee Aide, or make bill status and scheduling inquiries to Bernadette Kmetz, secretary at (609) 984-6843.

The Assembly Education Committee will meet on **Monday, April 19, 1993 at 9:30 AM, in Committee Room 8, Legislative Office Building, Trenton, New Jersey** to receive testimony from the Commissioner of Education, Mary Lee Fitzgerald, department staff and others concerning the department's skills testing program, including the early warning test and high school proficiency test.

Issued 03/26/93

ASSEMBLY RESOLUTION No. 113

STATE OF NEW JERSEY

INTRODUCED FEBRUARY 18, 1993

By Assemblyman ROCCO, Assemblywoman WEBER,
Assemblymen Wolfe, Charles, Doria, Gaffney,
Martin, Garcia and Nickles

AN ASSEMBLY RESOLUTION directing the Assembly Education Committee to investigate the skills testing program developed and administered to New Jersey school children by the State Department of Education.

WHEREAS, The New Jersey Department of Education has undertaken a skills testing program in order to raise academic standards in New Jersey public schools and improve student achievement; and

WHEREAS, The 8th-grade basic skills test, or Early Warning Test which is designed to identify those students who might have problems with the 11th-grade High School Proficiency Test has been identified as having serious problems; and

WHEREAS, The problems associated with the Early Warning Test may render the entire testing program, including the High School Proficiency Test, useless; and

WHEREAS, Reports have surfaced that certain Department of Education employees may have attempted to cover-up the problems with the Early Warning Test; now, therefore,

BE IT RESOLVED *by the General Assembly of the State of New Jersey*:

1. The General Assembly of the State of New Jersey, on behalf of the citizens of this State, for the purposes cited in the preamble hereto, directs the Assembly Education Committee to investigate the development and administration of the New Jersey Department of Education's skills testing program.

2. It shall be the duty of the committee to thoroughly review all aspects of the skills testing program and the conduct of State employees in regard thereto.

3. Within three months after this resolution is filed with the Secretary of State, the committee shall issue a report of its findings and make such recommendations to the General Assembly as it shall determine to be appropriate based upon an analysis of the facts resulting from the inquiry.

4. A duly authenticated copy of this resolution, signed by the Speaker of the General Assembly and attested by the Clerk thereof, shall be transmitted to the Chairman of the Assembly Education Committee.

STATEMENT

This resolution directs the Assembly Education Committee to review all aspects of the skills testing program developed and administered by the New Jersey Department of Education.

Allegations have been made that serious problems exist in the 8th-grade Early Warning Test which may implicate the entire skills testing program and which certain Department of Education employees may have attempted to cover up. The committee is required to report its findings to the Legislature within three months after the filing of this resolution.

Directs Assembly Education Committee to investigate the skills testing program developed and administered by the New Jersey Department of Education.

TABLE OF CONTENTS

	Page
Mary Lee Fitzgerald Commissioner New Jersey Department of Education	2
Leo Klagholz Assistant Commissioner Division of Academic Programs and Standards New Jersey Department of Education	5
Patricia Wang Iverson Private Citizen	68
Ellen Oppenheimer Private Citizen	68
Joseph G. Rosenstein, Ph.D. Mathematics Department Rutgers - The State University, and Director New Jersey Mathematics Coalition	79
Patricia Clark Kenschaft, Ph.D. Professor of Mathematics Montclair State College, and Director Project for Resourceful Instruction of Mathematics in the Elementary School - PRIMES	86
Ira Sweet, Ph.D. Teacher, Guidance Counselor, and School Psychologist in the New Jersey Public School System	90
APPENDIX:	
"Eighth-Grade Early Warning Test Findings and Recommendations" submitted by Commissioner Mary Lee Fitzgerald	1x
"Report on the Eighth-Grade Early Warning Text (EWT)" submitted by Commissioner Mary Lee Fitzgerald	12x

TABLE OF CONTENTS (continued)

APPENDIX (continued):

	Page
Letter addressed to The Honorable John Rocco from Thomas B. Corcoran	28x
Letter addressed to The Honorable John A. Rocco from former Commissioner New Jersey Department of Education, John Ellis	31x
Letter addressed to Assemblyman J. A. Rocco from Mrs. Diana Klarmann	34x
Brochures submitted by Patricia Wang Iverson	37x
Testimony and attachments submitted by Joseph G. Rosenstein, Ph.D.	40x
Testimony submitted by Patricia Clark Kenschaft, Ph.D.	44x
Testimony submitted by Ira Sweet, Ph.D.	46x
Letter plus attachments addressed to Alan J. Steinberg Assembly Majority Office from Louise Mead Riscalla, Ph.D.	52x

mjz: 1-46
hmw: 47-51
bgs: 52-78
mjz: 79-94

* * * * * * * * * * *

ASSEMBLYMAN JOHN A. ROCCO (Chairman): The meeting will come to order. David, will you call the roll, please?

MR. HESPE (Committee Aide): Assemblyman Garcia?

ASSEMBLYMAN GARCIA: Here.

MR. HESPE: Assemblyman Charles? (no response) Assemblyman Wolfe? (no response) Assemblywoman Weber? (no response) Assemblyman Nickles?

ASSEMBLYMAN NICKLES: Here.

MR. HESPE: Assemblywoman Anderson?

ASSEMBLYWOMAN ANDERSON: Here.

MR. HESPE: Assemblyman Martin? (no response) Chairman Rocco?

ASSEMBLYMAN ROCCO: Here. We have a quorum.

Obviously, we are all here for one reason, and that is to look at the testing program in the State; how we can improve that program to look at those things which may have gone wrong in the past; and to make certain that these problems do not reoccur in the future. We will attempt to take a look at some of the things that we have been notified about in regard to the test itself and how it was validated, or not validated. We will also, hopefully, have some recommendations from the Committee to the Commissioner in regard to the future of testing -- some possible recommendations by this Committee.

Before I start, maybe the Committee members have a comment or two. Fred, Priscilla, Rudy?

ASSEMBLYMAN NICKLES: Mr. Chairman, thank you. I appreciate the opportunity first to hear the Commissioner speak and give her testimony. I would like to reserve my comments until after that, please.

ASSEMBLYWOMAN ANDERSON: I, too, would like to hear the Commissioner first, and then I will have comments.

ASSEMBLYMAN GARCIA: I'll wait.

ASSEMBLYMAN ROCCO: Okay. This is a Committee hearing for Resolution No. 113, introduced on February 18. In that

regard, and without further ado -- and without dragging this thing out -- we will hear from the Commissioner in regard to some of the information she has available. Once again, this is nothing more than our attempt to try to resolve some of these issues and to do a better job for the children of the State. We are not here for partisan purposes.

Commissioner?

COMMISSIONER MARY LEE FITZGERALD: Thank you very much, Mr. Chairman.

There is a high level of growing dissatisfaction with the traditional means of assessing student progress in America. From all professional groups is emerging a call for broader measures that more adequately describe the degree to which students have learned what they have been taught.

If the efforts of school reform are going to effect lasting change, traditional forms of testing must yield to models of assessment that are embedded in the curriculum itself. Traditional testing measures the mastery of facts and the recall of information, both of which are easily formated into true/false and multiple choice items. The newer models of testing and assessment provide opportunities for students to explain what they know, and to apply the facts to prove it. Two- and three-step problems which require higher order thinking demonstrate real levels of understanding. The guess factor is minimized.

New Jersey's Assessment Program requiring an 11th grade graduation test to certify graduates have learned what the high school curriculum has taught, reflects these newer models of assessment. Students do not simply choose a multiple choice answer to indicate their understanding of subjects and predicates. They actually have to write a short essay in which the rules of grammar are demonstrated. Similarly, real problems expressed in mathematical symbols are described which require thoughtful solutions.

The class of 1995, this year's 10th graders, is the first class to take a graduation test at the 11th grade. This new graduation test was created by action of the Legislature on December 1, 1988, in order to continue New Jersey's incremental effort to raise high school graduation standards over the past decade. At the same time, the Legislature added the requirement that the Department of Education provide an early warning test for all 8th grade students for the purpose of providing information to parents, guidance counselors, and teachers regarding the students' relative strengths in mathematics, reading, and writing.

The current sophomore class took that early warning test in the first year of its administration when they were 8th graders in 1991. Their high schools have used the information, along with other achievement measures, to counsel the appropriate high school schedule. A year ago, the current freshman class took a second form of the EWT as 8th graders in 1992 in anticipation of the new graduation test in their junior year.

As we know, the efforts to make the 1992 test weight equal in difficulty with the 1991 test resulted in an extended dispute between CTB, the test contractor, and the Department of Ed. If you got a 50 in 1991, would you get a 50 in 1992, or were the items in 1992 slightly harder? If they were, perhaps a 50 in '91 would actually have meant the same thing as a 48 in '92. This statistical exercise is called "equating," and is regularly employed by the testing industry to make alternate forms of the same test the same level of difficulty.

During my first meeting with Luis Salgado in January 1993, who was then Assistant Commissioner for Educational Programs and Student Services, I was informed of a potential problem with the cut scores of the reading subtest of the '92 EWT related to the Department's attempts to equate the '91 and '92 tests. Because Dr. Salgado and Dr. Masonis, the Director

of the Department's Bureau of Statewide Assessment, suggested that students could have been inappropriately placed in remedial or supplemental reading classes based on our cut score, I brought the information to the State Board of Education at its monthly meeting on February 2. I indicated that I would undertake a thorough investigation of the matter and report our findings at the Board's March meeting on the 7th.

Dr. Leo Klagholz, now Assistant Commissioner for Academic Programs and Standards, conducted that investigation, with the assistance of Mr. Tom King, from our Compliance Unit. You have copies of the two reports presented to the State Board: one a comprehensive chronology of the circumstances surrounding the disputed cut score of the reading subtest, and the second a summary of our findings and recommendations. We have concluded the following:

The early warning test is not flawed. On that basis, I authorized the administration of the 1993 EWT in the public schools following the State Board meeting on March 7. All the school districts were advised in the printed guidelines: "Students whose scores fell below the standard on one or more sections of the EWT are candidates for supplemental instruction in reading, math, and/or writing. Final placement in supplemental instruction is to be determined after additional assessment information is considered."

Consequently, I determined that the great majority of school districts had used multiple measures to place students in these classes as advised, and had not used the single cut score.

Three, a single cut score should be replaced by performance bands to indicate the range of possible achievement in each of the subtests on the EWT, which, in fact, the Department had used in '91 but changed to a single cut score in 1992. This band of performance, as well as other assessments employed by each school district, should be used to assist

guidance counselors, parents, and the students themselves in planning an appropriate high school course of study.

I think you have a number four, and I will talk about that in questions and answers.

New Jersey's new 11th grade graduation test has been described as a cutting edge assessment which cannot be duplicated by an existing norm-referenced commercial test. It is successfully and effectively forcing an upgrade in our high school curriculum, specifically in our statewide mathematics curriculum, which addresses the new standards recommended by the National Council of Teachers of Mathematics. Its writing subtest has forced schools to incorporate process writing in the upper elementary and middle grades, and hundreds of English teachers are now trained in holistic scoring. The reading test contains sophisticated passages of complex information, some of it specifically referencing New Jersey, requiring students to analyze information and apply solutions to problems.

The dozens of core proficiencies developed by panels of high school teachers over the past five years form the basis of this graduation test. Curriculum has been realigned; new textbooks have been purchased; and new programs have been implemented. Slowly and consistently over the past 10 years, the content of the high school curriculum in New Jersey has been adjusted up. Standards are higher, students more serious, and the scores more academically focused. Both the 8th grade EWT and the 11th grade HSPT are contributing to this renewed rigor in the curriculum, which is strengthening the integrity of New Jersey's high school diploma.

I am joined today, albeit late--

ASST. COMM. LEO KLAGHOLZ: I apologize.

COMMISSIONER FITZGERALD: --we came separately -- by Dr. Leo Klagholz, who is the Assistant Commissioner for Academic Programs and Standards, and in back of me, Dr. Edward Masonis, who is the Director of Statewide Assessment in the Department. All of us will answer any questions you may have.

ASSEMBLYMAN ROCCO: Our Vice-Chair has joined us -- Assemblyman Martin.

There are a number of questions I think all of us have. Some you will probably be able to respond to, and others you may not. Just for the record, there are letters in from John Ellis and Tom Corcoran that we will maybe refer to later, but which certainly will go into the record with regard to their perspective with regard to the testing.

I guess what bothers me the most, and others I have spoken with in the Legislature, and Assemblyman Doria, who will follow-- I think for well over a decade we have been concerned about the ability of the State to develop these tests and to provide our young people with an instrument basically that will do the job in terms of evaluating the early warning issue and will do the job in terms of determining whether or not they have the proficiencies to graduate. I think what has bothered me -- and I will let Joe speak for himself -- what has bothered me over this period of time, has been that the Department has just had this drive to develop their -- "their tests," when in fact, I think clearly we see that the Department does not have the capability or the competence or the time or the money to truly do that sort of a job.

At some point, I would like the Committee -- who all might not be involved in education in terms of it being part of their professional lives -- to maybe get a quick review of item analysis, validation, how a test is truly put together, and how it is structured to make certain that it consistently serves the purpose that it is intended to serve.

So in brief, number one, I guess the first question is: How much are we spending per year -- this State -- to develop tests that are already out there in the market? I guess that would be the first question.

COMMISSIONER FITZGERALD: The Department's testing budget has run roughly $2 million a year to develop a New

Jersey graduation test, which the Department has not thought was available commercially on the market. One of the problems with the ability of the Department to do the validity studies and the field studies that, in retrospect, should have taken place resulted from the fact that the budget was significantly cut.

Obviously, we have been doing a lot of data gathering on this topic, and we have gotten a lot of information from other states; the states that do have graduation tests and the amount of money they have been spending on doing what New Jersey has been doing. Actually, relatively speaking, we are not spending a lot of money developing a graduation test for New Jersey's high school students.

ASSEMBLYMAN ROCCO: There were other years, I think, where we spent significantly more than that. So you're saying that $2 million is about an average?

COMMISSIONER FITZGERALD: Yes. This year we have a $4 million request in, however, so that's not too--

ASSEMBLYMAN ROCCO: In retrospect, you're indicating clearly that the field testing and the validation studies were not done properly?

COMMISSIONER FITZGERALD: Sufficiently, I think. There were not enough of them.

ASSEMBLYMAN ROCCO: Properly, sufficiently?

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN ROCCO: Which, once again, the State being the volatile entity that it is, one year you could be doing very well financially and the next year, as we have seen, we can be in a depressed situation where moneys may not be available and we may not be able to pool the resources and do the job. But, suffice to say, it was not done sufficiently or properly. So therefore, we came up with data that was inaccurate. And in that process, many students -- 5000 that we have identified -- have been misplaced. Lord knows what other kinds of problems have occurred over the years.

I do not represent Bayonne, but I did get a letter from a Diane Klarmann from Bayonne, which basically represents, I think, many, many parents in the State of New Jersey. We will certainly put this in the record, but it clearly indicates that because this student, who happened to be an honor student, did not do well in the early warning test, they were put into remediation for one semester, and maybe more, if the institution -- the school is not advised that there were errors, and if she was one of those people -- or the child was one of those people who was misscored.

I know you are indicating now that we are indicating to principals and superintendents that it is not a definite line on the test scores one way or the other; you know, that there should be flexibility within the scores before a placement occurs within remedial courses. But apparently many principals and superintendents around the State felt the need that if they didn't pass the early warning test or reach a certain score, that they had to be placed in remediation. So therefore, 5000 or so identified have been placed in remediation, and many others through the Guidance Department, etc., if we don't straighten this out, may have greater difficulty.

I guess the first question I have in that regard, because it ties into the lack of validity in the previous question, is: Have these parents and students been notified that they were misplaced?

COMMISSIONER FITZGERALD: Well, first of all, there is a question as to whether or not they were misplaced. We haven't finally determined that they were misplaced. We have several memos that superintendents received. I received one as a superintendent, I have to say, every year that these tests have been coming out, which said: You should not automatically put students in remedial classes if they fall below a cut score, because teacher judgments, grade level placements, grade

point averages, consider a lot of other things-- We ought never to use a single cut score -- I think all of us understand that -- when we are making educational decisions about kids.

So, they had been notified. There were even regional meetings last fall set up with people -- representatives from school systems -- going over, as a matter of fact -- which is delineated in one of those reports you received -- explaining the flexibility and interpreting that cut score. The first year there was a band; there wasn't a single score. And the second year -- which of course contributed to a lot of the controversy -- they decided to set a cut score.

We have, yes, talked to the superintendents. I have sent information to the superintendents suggesting that this is a local decision; what they should do, if anything. The superintendents have informed me that that is what they have done; handled students on an individual basis. Many of them have said to me that they, on their own, put more students into supplemental instruction than a cut score would indicate, because they want to err on the side of the student, and students definitely would profit from additional help if they were close to any line or performance band.

It isn't that the test, in and of itself, the 8th grade warning test, was flawed and that the data that we got from that test, in any way, was wrong. In discussing whether or not to -- where to set this performance band, or cut score, to make it the same in '91 as it was in '92, was really the only issue which surfaced and has created such an intensive review of the test itself, but a lot of people have spent a lot of time, and not just people in our own Department. I think, Mr. Chairman, the more I have looked into this, I, too, wondered why should the Department of Education develop a test, when there is a very big, sophisticated industry out there of test makers. We certainly do not have the capability within our own staff, but we have contracted out lots -- parts of this

test. We have not done it all oursleves, which is part of the problem. One of the people we contracted it to obviously had some disputes with us related to equating, but we have not created these tests entirely on our own during the last decade, but have used, in most instances, testing companies to assist us in their development.

ASSEMBLYMAN ROCCO: Before I go to Assemblyman Nickles, who is also a Superintendent-- I have had a few courses in tests and measurements, statistics, etc., but I do not claim to be an expert by any stretch of the imagination in that regard. But, in the item analysis, as you went through each test question, or the Department did-- At that time, you were not the Commissioner, obviously. Were there test questions that were thrown out after the test was given?

COMMISSIONER FITZGERALD: In '91 and '92, yes.

ASSEMBLYMAN ROCCO: What was the purpose of throwing those test questions out?

COMMISSIONER FITZGERALD: To make the test a better test, more valid, which happens-- That is one way to conduct validity studies. When you are field testing a test, to see those items-- An unexplainable number of students either pass it or fail it. Then those test items are removed.

ASSEMBLYMAN ROCCO: They are considered unstable, right, basically? But these are thrown out after the students received their evaluation, whether it was 50, 60, 70, 80. I mean, the individual student had that number recorded, and then the question was thrown out afterwards--

COMMISSIONER FITZGERALD: No.

ASSEMBLYMAN ROCCO: --or was it thrown out before the district received the individual test scores?

COMMISSIONER FITZGERALD: Dr. Klagholz?

ASSISTANT COMMISSIONER KLAGHOLZ: They were thrown out before scores were assigned to the students, and there was a very minimal number. We were lucky on that. If we had done

field testing, which would be to administer the tests in advance to a sample of students to see how they would do and throw the questions out then, and then administer it, that would be one thing. Instead--

ASSEMBLYMAN ROCCO: So you did the item analysis. You threw the question out before--

ASSISTANT COMMISSIONER KLAGHOLZ: And then scored it.

ASSEMBLYMAN ROCCO: --you graded the papers, so to speak?

ASSISTANT COMMISSIONER KLAGHOLZ: In fact, they were thrown out after the test was administered, but thrown out from the scoring process.

ASSEMBLYMAN ROCCO: Did it deal with the fact that many students may have failed that individual, specific item?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes, that's it. Yes. It could be any number of variables. It is whether or not the item works. Are the good scorers getting it right and the low general scorers getting it wrong? Or, is everyone--

ASSEMBLYMAN ROCCO: Okay. I guess the concern in that sequence is that you throw out the difficult items to have a higher test score, so that it looks like there is an increase from one year over the next.

ASSISTANT COMMISSIONER KLAGHOLZ: Actually, as it turned out, there were a few -- a very few items that had to be thrown out. It was two, Assemblyman, I believe. I think the point you are making there is exactly the nub of the situation; that is, the first year the test was valid, but it was the first year and it wasn't an ideal version. There was an attempt the second year to equate the cutoff score to the first year, which was unstable, and it should not have been done.

ASSEMBLYMAN ROCCO: Okay. Go ahead, Fred.

ASSEMBLYMAN NICKLES: Thank you, Mr. Chairman. A statement, and then a couple of questions. I would like to react, first of all, concerning the 5000 students. That was a

question. It is true that every school district, especially since the problem with this testing arose-- It is the understanding that you never use-- You always want to use multiple standards. However, if I had to assign a percentage weight to how much importance you would place on the testing, I would say, as a Superintendent, probably in the 90th percentile. The reason for that is that the liability you have as a school district, if a student was not placed into basic skills or a remediation program, and somewhere down the road a question arose from monitoring, and they were under the MLP, or the minimal liberal proficiency, or the cut score, as we refer to it, then I think the district would have a liability to address.

So, yes, school districts are required -- should use multiple standards, but when you have a documented, valid, and reliable test to work off of, that probably weighs 75 percent to 90 percent of your decision-making process.

COMMISSIONER FITZGERALD: Sure, I understand that. We did, too, in Montclair. Sure.

ASSEMBLYMAN NICKLES: I just wanted to put that out for the Committee.

ASSEMBLYMAN ROCCO: Fred, let me just elaborate on that a little. From the feedback I have received -- and I think other members of this Committee have received, and I assume Assemblyman Doria, as well -- the schools-- It is kind of like, you know, taking the SAT scores, when the Admissions Committee says, "Well, that is not the only factor we use." But we know you don't get into school "X" unless you have a SAT of a certain given number.

I think what happened down in the rank and file principals and superintendents, is that they do, in fact, use that test score for placement, and they do use it as a marker, because it is the most-- Well, it is the easiest marker to use. I mean, it's, you know-- Those above don't have to go

into remediation; those below do. I think a great deal of that occurred from the letters I received and from the memos I received in that regard. So I think that did happen. Whether, you know, the Commissioner at the time, John Ellis, specified or not-- Apparently he did specify not to use it as the only criteria, but my guess is that in many districts that really wasn't followed; that principals and superintendents kind of just used that marker.

Fred?

ASSEMBLYMAN NICKLES: Thank you, Mr. Chairman. Also, I would like to preface my remarks by saying that neither the Commissioner nor the Assistant Commissioner nor I were here when all of this began.

ASSEMBLYMAN ROCCO: That's true.

ASSEMBLYMAN NICKLES: So the remarks are not directed at any individual, but rather at the problem that exists.

I guess I would like to ask the Commissioner, or the Assistant Commissioner-- For example, in 1990, the students in the K through 12 school districts were receiving State-mandated tests at what levels, and how would they differ, say, in 1995?

COMMISSIONER FITZGERALD: Ask that again.

ASSEMBLYMAN NICKLES: Can you compare 1990-- What grade level schemes would be required, State testing, and what will the requirements be in 1995, for example?

COMMISSIONER FITZGERALD: In 1990, you were required to annually assess students in grades K through 12. You only reported your scores with a MLP in grades three and six to the State. In grades K through two, there was not a MLP, so your own district created their own assessments. The 19-- I am not sure if this was 1990. I am trying to remember when this happened because I was a Superintendent. Let us say, today -- I think it is better to say today-- Today, you are not double testing at the eighth or 11th grade, but we were until very recently. The 8th grade EWT now can take the place of the

State's requirement of an annual assessment. It was-- We had to also give a standardized test as well as the 8th grade test, and you do not have to give a second standardized test in the 11th grade. The 11th grade HSPT now is the required annual testing for 11th graders. That is the State requirement.

What happens, however, is that there are Federal requirements that end up requiring double testing, primarily for Chapter I students. If you have Chapter I students, they will require a second measure and a norm reference test at grade eight and grade 11. The State does not technically require it, but to receive and use Federal moneys in Chapter I does. However, the reauthorization of Chapter I is suggesting that even Chapter I is going to become much more flexible in replacing the traditional norm reference test with something else. So I don't think that is going to continue another year.

ASSEMBLYMAN NICKLES: Okay. So, we are basically looking, as far as the State is concerned-- Are we not looking in the future to a fourth, eighth, and 11th grade testing?

COMMISSIONER FITZGERALD: Four and eight, rather than three and six.

ASSEMBLYMAN NICKLES: Okay. Of course, as general information, the HSPT test came out and it was given in 9th grade, and then the test was upgraded to a higher version which was more thinking assessment oriented. Now we are moving from that higher level thinking assessment test in the 9th grade to the 11th grade.

I guess I would like the Commissioner, or her Assistant, to explain the purpose for moving it to the 11th grade, the philosophy behind it, and also, if you could perhaps address for the Committee the concerns that many of us have received, particularly from the urban communities that are fighting the problems, as the special 30 school districts, and the potential takeover of school districts. Is it really fair for those school districts -- particularly those, and others as

well -- to do the remediation necessary, understanding that identification will be made of schools in 8th grade, but in order to get that high school diploma, when the test is taken at the 11th grade, the remediation of that is still limited as far as the time period is concerned?

Perhaps, would we be prudent to rethink giving that test and compromising, as the Chairman has suggested, perhaps in the 10th grade, with a couple of years to do the remediation? Now, there is a lot there, and I will be glad to go over it. You don't really have to write all of that down.

COMMISSIONER FITZGERALD: I think the remediation issue is what led the legislation to suggest an early warning test in the 8th grade, so that instead of remediating, we could begin, proactively, to more intensively made sure the students had the skills to pass the test in the 11th grade.

It is a harder test, Assemblyman. It is a harder test. I mean, I was involved and looked at some of that myself as a Superintendent. I was told -- and I participated in some of the discussions as a Superintendent -- that we were incrementally raising standards in the State of New Jersey. Moving from a very dumbed down basic skills test, we took a step. We created a harder test in the 9th grade. Since this is a graduation test, it made some sense that we weren't going to accept 9th grade skills for a high school diploma, but we were trying to get closer to what the students were learning as in the full high school curriculum.

Certainly the State Department of Education, as well as all of us, has been criticized in the last decade about not having sufficient standards to meet the challenges of the 21st century. It was an intentional decision to make a harder test. I don't think anybody would argue that in the Department, nor any of the people who were involved in raising those standards. It definitely is a harder test.

The issue of the urbans-- We certainly know that the students in the urban schools have not been performing as well on the 9th grade test as students in nonurban areas. Again, using the 8th grade test as an early warning, I think it again puts a lot of responsibility on people in the upper elementary grades, which is another reason that it is a good idea to have an 8th grade test. High schools obviously cannot do this themselves; start saying to elementary schools, "You really have to make sure that you meet your responsibilities. These kids start getting behind." But I am the first one to feel a little overwhelmed with the lack of achievement we are getting out of most of the schools in urban areas, and knowing that these kids have to be just as productive as kids coming out of suburban schools, or we certainly have not met our obligation to their future.

ASSEMBLYMAN MARTIN: Fred, may I--

ASSEMBLYMAN NICKLES: Sure.

ASSEMBLYMAN MARTIN: May I ask a question that is related to that, to help me out?

ASSEMBLYMAN ROCCO: Assemblyman Martin, our Vice-Chair.

ASSEMBLYMAN MARTIN: What do you anticipate-- Suppose a student does not do well on the 8th grade-- By the way, I took these tests, at least in the booklet you sent out. I found the 11th grade test to be pretty difficult. I mean, maybe that's--

ASSEMBLYMAN ROCCO: But he is an attorney, and you can understand that. (laughter) Strike that from the record.

ASSEMBLYMAN MARTIN: I'm not shy, but I didn't ask to be graded on it. I just took it. I also looked at the 8th grade test.

Let's say a student does not do well on the 8th grade test. What do you anticipate, as Commissioner, that a student would be doing from 8th grade to 11th grade in addition to what they also have to do with their 8th grade graduation, or high

school curriculum, in order to be able to pass? I am not quite clear how someone would be able to gather that remediation and also perform their other responsibilities in high school. What kind of a program would be available to that student?

COMMISSIONER FITZGERALD: Well, I can tell you what we were doing in Montclair, because I just came from that school system. We had summer programs that we specifically set up -- intensive summer programs. If you graduated from the 8th grade and you were even in the middle, we wrote a letter to the parents encouraging them to take advantage of the opportunity to spend six weeks in an intensive catch-up.

When you start scheduling kids into most high school programs with seven or eight periods, you certainly can influence their electives; what they are choosing to take. But I think beyond that what this test is doing, is forcing all of the departments -- all the departments, social studies and science -- because, let's face it, the reading section-- You have to read materials which are technical, materials that you read in a science class or a social studies class. I think it is making all teachers, or it should in public high schools -- teachers of reading and writing and teachers of mathematics-- The teachers in an English or mathematics class cannot possibly be responsible for creating a total program for these kids. So, the departments themselves, I believe, are taking additional responsibility and making sure that what they are teaching in all of their courses are: a high order of thinking skills, and problem-solving applications that go far beyond simply answering the questions at the end of a chapter.

A lot of schools, though, are doing a lot of other things. In study halls they are setting up mentors, peer coaching. This is a very serious obligation of ours to level up, and not simply be content with what kids have done in the past. I think parents and people in the community, certainly the people in the business roundtable that I have been meeting

with off and on since I came down here to Trenton, are telling me, "Kids are not working hard enough. Schools are not requiring enough, and we have to recreate a strong academic focus in these public high schools if these kids are going to be productive."

ASSEMBLYMAN MARTIN: I know I am cutting into Fred's time, but--

ASSEMBLYMAN NICKLES: Let's hear it.

ASSEMBLYMAN MARTIN: I have two related questions: One, in your view, is it working, this remediation? And secondly, John has always been concerned about whether there is going to be an undue emphasis on people -- teachers teaching for the test, if you understand what I'm-- You know, working just to try to get the test passed, rather than on education in general. Maybe you can answer those questions.

COMMISSIONER FITZGERALD: Well, they are supposed to teach to this test. This is a high school graduation test, and it is supposed to reflect what they have learned. So they really are supposed to be teaching the material that is on the test.

ASSEMBLYMAN MARTIN: So, in your view, teaching to this test gives you enough of a broad-based educational achievement level; that that is okay?

COMMISSIONER FITZGERALD: Basically, it is sort of like a criteria reference test. You are teaching the content that is going to be tested. That's correct.

ASSEMBLYMAN MARTIN: Is it working, not the test, but going back for this remediation? You said that if students are performing not satisfactorily in 8th grade, for whatever reason, there is an early warning signal that goes off. Are those students going to be-- Do you find that with this additional study hall and mentoring and asking them to come to summer school, and so forth-- Are those things working for most of the students who have this difficulty?

COMMISSIONER FITZGERALD: Of course, next year is the first year that these kids are going to take it as a high school graduation requirement.

ASSEMBLYMAN MARTIN: But we have had the early warning program.

COMMISSIONER FITZGERALD: For two years, that's right. The kids right now who are sophomores were the first ones who took the EWT.

ASSEMBLYMAN MARTIN: I know. I have a daughter who is a sophomore, and one in 8th grade, so I know that they have gone all through that.

COMMISSIONER FITZGERALD: So, you know. Some of you know that I got married last fall. I married into a family, unfortunately for me, with a sophomore in the local public high school, so I have been hearing all about this as well.

I don't have any hard data. All I have is from talking informally to department chairs in math and English in the public high schools. They say how focused they are on meeting this commitment.

ASSEMBLYMAN MARTIN: So you would hope, and maybe assume, but you can't tell us with any degree of surety that those -- that the number of students who had a problem in the early warning system in 8th grade, that that percentage will decrease when they actually take the HSPT in the 11th grade?

COMMISSIONER FITZGERALD: Well, if I had been here two years ago, we might have set up some way, see, to gather this data. Are there fewer kids-- I mean, look at their grades in English and math. There could have been ways, although this is a mammoth data gathering exercise. I simply don't know that yet. I would assume high school principals have that data; know how many kids each year are required in Chapter I for remediation -- and they were using the EWT as an indication of that -- but I don't have the data.

I just happened to think that there is something else I would mention to you, Assemblyman. We just got notice that an additional $8 million is coming into New Jersey this year, compliments of Chapter I down in Washington. We have a meeting this Friday with the eligible districts -- most of them are urban; we have the amounts of money that are going into their districts -- to describe for them our guidelines for the use of that money. A lot of it, as you might imagine, at the secondary level, we are targeting to issues of the HSPT-11 and the EWTs. That is an additional allocation of money that they were not anticipating. I am sure they are going to take advantage of it this summer.

ASSEMBLYMAN NICKLES: Mr. Chairman?

ASSEMBLYMAN ROCCO: Fred is going to finish up; then Joe Charles; then David. But before we do that sequence, the former Chairman of this Committee has joined us -- Joe Doria -- who has also had an interest in this issue for well over a decade. We are happy to have you here, Joe.

ASSEMBLYMAN DORIA: Thank you, Mr. Chairman.

ASSEMBLYMAN ROCCO: Is there anything you want to say, Joe?

ASSEMBLYMAN DORIA: No, I'll wait.

ASSEMBLYMAN ROCCO: Okay. All right, we will have Fred, Joe, and David. Go ahead.

ASSEMBLYMAN NICKLES: I just want to make a final comment: I have every confidence that the teachers and the administrators throughout the State of New Jersey can implement higher standards. My concern, and I say this knowing that tomorrow every type district in the State of New Jersey will be voting on school budgets and refinancing and everything else that goes on tomorrow-- I have a concern, as a Superintendent as well as an elected State official, that there is a cloud that surrounds our testing program right now.

It seems that the media and opponents of education many times point out the negatives and use that to confound, or convolute, what the local school districts are doing. My concern with the whole testing program, and a cautious approach that I have, is that this somehow cannot be misused throughout the State of New Jersey to show that our students are doing poorer than in previous years; that the tests aren't as responsible as they should be. I really want to make sure myself that our testing program is above reproach and that it is a positive reinforcement of what the students are learning.

So, my cautiousness is based on what the public reaction could be, may be, due to a problem that has existed for a number of months.

I want to thank the Commissioner and the Assistant Commissioner for their comments.

ASSEMBLYMAN ROCCO: Fred, I'm sure you have many questions to follow. I think Fred's issue there is significant, in that if our test is not validated properly and we are spending all of this money to show inaccurate data possibly, then we truly have to wonder whether or not we have used the right instrument, and truly whether or not it is testing what it is supposed to test. I will come back to that.

Joe?

ASSEMBLYMAN CHARLES: As a follow-up on Assemblyman Martin's question, Commissioner, where the remediation is concerned, test scores are given, problems are recognized, and then you described what happened in the School District of Montclair. That is in Montclair. The other districts throughout the State of New Jersey-- Are there guidelines that come forward from the Department of Education with respect to what the school districts should do? Are there guidelines or mandates about what has to be done with regard to remediation; like, for example, hold summer programs, and guidelines on who must attend, and so on? Or is that something that is just helter-skelter throughout the State of New Jersey?

COMMISSIONER FITZGERALD: No, there are definitely guidelines. Beyond guidelines, there is technical assistance; there are meetings; there are workshops with people in districts. I mean, we created ours based on some meetings we had attended.

ASSEMBLYMAN CHARLES: I guess my question specifically is: Are there concrete programs that every school district must provide to address the remediation problem that is recognized in these early warning tests?

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN CHARLES: Are those the summer intensive programs that you described? Is that something that every school district has to have?

COMMISSIONER FITZGERALD: It depends upon whether or not-- Most of these districts are Chapter I eligible, and most of their additional money for remediation comes from Chapter I, which is a Federal remedial program. There are definite guidelines. In fact, there are requirements as to how you use Chapter I money: how many kids, how long, all of that. All districts eligible for Chapter I money are very conversant with what those requirements are.

ASSEMBLYMAN CHARLES: How many districts out of the number of districts in the State of New Jersey-- What percentage, roughly, are Chapter I eligible? All of them? Which ones? That goes to my question about where the programs are. You answered my question in terms of Chapter I districts. Now, that gets back to my first question about the school districts -- all of them throughout the State of New Jersey. If half of the districts are Chapter I eligible, that leaves the other half that are not Chapter I eligible. What happens to those districts?

COMMISSIONER FITZGERALD: Right. Until two years ago, the State provided money under State Compensatory Education with the same kinds of guidelines and criteria for entrance

into those programs, as well as, you know, defining the description of what they were going to be. With the QEA legislation, State Compensatory Education has been renamed. It is now called At Risk Aid. There is additional money for students who are academically at risk and, yes, there are suggestions and guidelines from the State Department related to those students.

There are very few districts in the State of New Jersey which do not qualify for one or two of those programs. In the several that do not, there are still monthly meetings with all school superintendents, with the County Superintendent of Schools. At those meetings, the Department gives out information, suggestions, communiques about any of these issues.

ASSEMBLYMAN CHARLES: Okay, let me-- I guess my question to tie that subject matter down is: The children who are found to be in need of the remediation--

COMMISSIONER FITZGERALD: Yes?

ASSEMBLYMAN CHARLES: --what percentage of them actually gets and undergoes remediation, as opposed to the programs which are in place? Do we have a requirement that each kid who has a problem has to have remediation? Do we find that only 50 percent of those who are identified as having a need for remediation are actually taking advantage of those courses or programs which are available? Or, do we find that a larger percentage of all of them are?

I have a concern that, you know, we have meetings of administrators and everybody discusses what the problem is. Meanwhile the kid who doesn't know anything, who needs help, doesn't get the program. Is that a legitimate concern?

COMMISSIONER FITZGERALD: No. I would say that except in very unusual circumstances, all kids in the State of New Jersey who qualify for remedial education, are getting it.

ASSEMBLYMAN CHARLES: Then one question I'm--

ASSEMBLYMAN ROCCO: That's a pretty broad statement. I hope it is true.

ASSEMBLYMAN CHARLES: I am willing to explore that, but I will accept it for the moment, Mr. Chairman.

My other question: This number of 5000 that we have been talking about as those who have been mislabeled, I guess, those who have--

COMMISSIONER FITZGERALD: Misplaced.

ASSEMBLYMAN CHARLES: --been misplaced-- Where do they come from? Where are they? In what districts? Are they in the urban 30, or--

COMMISSIONER FITZGERALD: They are from all districts.

ASSEMBLYMAN CHARLES: Do you find a preponderance of -- a disproportionate number in any type of district, whether it is suburban, rural, or urban?

COMMISSIONER FITZGERALD: There is a disproportionate number of them coming out of the urban special needs districts because there is a disproportionate number of underachieving students in the 30 special needs districts.

ASSEMBLYMAN CHARLES: So, have we developed data to know what percentage of that 5000 comes out of, let's say, the urban 30? Do we know that, or is that something that can be discussed?

ASSISTANT COMMISSIONER KLAGHOLZ: We haven't analyzed it, but we--

COMMISSIONER FITZGERALD: We could get that for you.

ASSEMBLYMAN ROCCO: You certainly would be entitled to that, Joe.

ASSEMBLYMAN CHARLES: Is there an estimate -- a good estimate -- for you to give right now as to-- Of these 5000, are we talking about 50 percent of those, or are we talking about 75 percent? Somebody must have some idea about that.

COMMISSIONER FITZGERALD: I think it would be dangerous to hazard a guess. We don't really have it, but we would be happy to get it for you.

ASSEMBLYMAN CHARLES: We have been talking about 5000 and we don't know where they come from? I just don't understand that. I mean, that is the first question that comes to my mind. We're talking about 5000 kids. Where are they? Who are they? What districts are the most affected? You know, I am just a little surprised that we don't know specifically, right here and now, since it has been such a topical subject, where they are.

COMMISSIONER FITZGERALD: We know the kids--

ASSEMBLYMAN ROCCO: I would tend to agree with Assemblyman Charles in that regard.

COMMISSIONER FITZGERALD: We know the kids and we know the districts. We don't have that information here today, so we will call back and get it for you.

ASSEMBLYMAN ROCCO: Are you finished, Mr. Charles?

ASSEMBLYMAN CHARLES: Yes, thank you.

ASSEMBLYMAN WOLFE: Commissioner, thank you for coming. I apologize for being late, like a student. I was actually here, my books were here, but I was out of the room when you were here. So, please give me credit.

I believe you have sensed from the tenor of the questions that this is not a partisan issue, but something that I think we share equally both as Republicans and Democrats. For my own confusion, and perhaps edification, the minimum basic skills test is required for 9th graders. Correct?

COMMISSIONER FITZGERALD: Currently.

ASSEMBLYMAN WOLFE: They must pass that to graduate?

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN WOLFE: Okay. The High School Proficiency Test is designed for 11th graders.

COMMISSIONER FITZGERALD: The 9th grade test is being phased out. That is correct.

ASSEMBLYMAN WOLFE: The 11th grade test?

COMMISSIONER FITZGERALD: Is just starting. Next year it counts for the first time.

ASSEMBLYMAN WOLFE: And the early warning test is for 8th graders?

COMMISSIONER FITZGERALD: Right.

ASSEMBLYMAN WOLFE: Now, in the information we are given -- this is a March 1993 report by Thomas King -- on pages 4 and 5 it describes these tests. On page 4 it indicates that a student must pass the minimum basic skills test in order to graduate. My question to you -- not in an adversarial way, but in a very concerned way -- is: Is this actually the case in New Jersey?

COMMISSIONER FITZGERALD: That you must pass the 9th grade test to graduate?

ASSEMBLYMAN WOLFE: Yes.

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN WOLFE: All students who graduate from schools in New Jersey have passed the 9th grade test?

COMMISSIONER FITZGERALD: There is a procedure, an SRA procedure, a student individual review that will allow a student who has not, between the 9th grade and the 11th grade, passed a test, to be individually assessed to see if he just can't take a test, but still knows the information.

ASSEMBLYMAN WOLFE: Okay. Now, what happens to the student who is in that situation when it comes time to graduate?

COMMISSIONER FITZGERALD: It is possible to graduate without passing the test.

ASSEMBLYMAN WOLFE: And what does his diploma indicate when he graduates?

COMMISSIONER FITZGERALD: It is the same diploma.

ASSEMBLYMAN WOLFE: So there is no way that a college or a perspective employer--

COMMISSIONER FITZGERALD: No.

ASSEMBLYMAN WOLFE: --would have any knowledge whether the student he would bring into the institution or the business has actually met the criteria for graduation?

COMMISSIONER FITZGERALD: Well, it is the same criteria, but it isn't measured on the test. As some of us know, some of us have experienced the fact that we know a lot more than we can, for some reason, indicate on a test. It is very rigorous. It is done through the County Superintendent's Office.

ASSEMBLYMAN WOLFE: Good. I share that 100 percent with you. We had a meeting with the Board, I believe it was two months ago when I believe you had your retreat. I made the statement to those people there that I was concerned that the test which was then under discussion, which was the test that we are talking about today, was being publicized in many regions of the State. The schools were gearing up for the test. Teachers were teaching for the test. You just indicated previously that in preparation for the test students were actually being taught geared toward the test.

I don't personally believe that the test measures what everybody knows. I think there are obviously other ways to do that. So I think that is my first concern; that our schools are teaching to the test. I am concerned as a legislator and as an educator in a college about accountability, but also about excellence. I will grant that not every child should nor necessarily wants to go to college, or even to a vocational school, or perhaps to some other level. But if a student is certified as meeting the graduation requirements, then certainly I think it should indicate that there are certain minimal skills that that child, or that young adult does possess.

Chancellor Goldberg has suggested that the schools from which the students graduate which are not proficient, be charged for the costs of educating those students while they

BEST COPY AVAILABLE

are enrolled in college. My concern is, if we are just letting people out, giving them a diploma and they are not really educated, regardless of what criteria we are establishing as a graduated citizen, or an educated citizen, then I think we are fooling ourselves and we are fooling the students and we are fooling the electorate.

So I would hope that as a result of this discussion we are having today that we can come to some consensus, whether it is the State's tests or a commercially derived test, you know, what are the standards that the residents of our State should expect that the graduates of the high schools should be able to meet? If a student does not meet those criteria, what, if anything, should be done to indicate that that student has not met those criteria? I am not saying it as a stigma, but I am saying it as a way to help that individual.

ASSEMBLYMAN ROCCO: David, I think to further clarify that, next year you are going to have the 11th grade proficiency, at which time if they do not pass that test, they will not receive a degree -- a high school degree -- with the normal diploma situation, but only an attendance degree. Is that correct?

COMMISSIONER FITZGERALD: No. It is the same diploma. It is the same diploma and the same procedure exists for the 11th grade test as on the 9th grade test. There is a SRA review for students who are seniors and, for whatever reasons, are bad test takers, but who still have met the same criteria, but demonstrated it in another way.

ASSEMBLYMAN ROCCO: But let's say the student flat out flunks next year's 11th grade proficiency exam -- okay? -- a SRA review just shows that they have not worked up to where they should be in terms of being able to pass that exam. They are going to graduate anyway?

COMMISSIONER FITZGERALD: No.

ASSEMBLYMAN ROCCO: I think that is your question, right, David?

COMMISSIONER FITZGERALD: They are denied a State diploma.

ASSEMBLYMAN ROCCO: They are denied a State diploma if they do not go through the SRA review--

COMMISSIONER FITZGERALD: Successfully.

ASSEMBLYMAN ROCCO: --successfully, to indicate that they are either test phobic or have other problems.

COMMISSIONER FITZGERALD: That is correct.

ASSEMBLYMAN ROCCO: I think the answer is, if they do not pass the high school proficiency exam in 11th grade, they do not graduate -- period.

ASSEMBLYMAN WOLFE: Right. I am glad to hear that, because my concern is, number one, for the child, but also for the teacher; the teacher who is under pressure either from the parent or from the board or from the district, to pass a student who is not prepared. My concern, therefore, has extended to the point of: What is the State of New Jersey, the Department of Education, prepared to do for that student who is going to be issued that special diploma? Are we going to be doing other things for him to try to help him, or is that it?

COMMISSIONER FITZGERALD: Assemblyman, I don't quite understand what you mean, a special diploma. A non-State diploma?

ASSEMBLYMAN ROCCO: An attendance diploma versus the graduate.

ASSEMBLYMAN NICKLES: There is a concept -- if I may, Mr. Chairman -- that if you are not a classified student, if you don't go through the SRA process--

ASSEMBLYMAN ROCCO: And you fail the exam.

ASSEMBLYMAN NICKLES: --and you fail the exam, do you get anything, you know, upon graduation. At one time there was talk about a certificate of attendance, that you attended.

ASSEMBLYMAN ROCCO: Yes. I think the certificate of attendance is the one that we talked about most often.

COMMISSIONER FITZGERALD: Yes, but it doesn't exist. It doesn't exist.

ASSEMBLYMAN WOLFE: That is all I really want to ask right now. I want to thank you for coming, and again just reiterate, I am not saying this in an antagonistic way. I am saying this in the sense of doing the very best we can, regardless of who it is or where they live or what district they are from. If a child graduates from a New Jersey high school, it will be a diploma of excellence.

COMMISSIONER FITZGERALD: I absolutely share that concern.

ASSEMBLYMAN ROCCO: Thank you, David.

Priscilla Anderson who has joined us, not normally a member of this Committee, brought to my attention some of her concerns as a guidance counselor in the Trenton School District. Priscilla has a few questions, we well.

ASSEMBLYWOMAN ANDERSON: Yes, Commissioner. I am always happy to see you here.

I understand Assemblyman Wolfe's concern as an educator at the higher level. I am an educator in one of the 30 special needs districts. I know that the majority of our youngsters have the ability, but because of so many social factors they are not achieving. I am concerned about that. We have such a problem in this country because of some of the problems which we are all aware of.

I am concerned that the test is-- I know that when people hear me say this, people right away say, "Oh, there they go again." But I do believe, from what I have seen, and from talking to so many teachers within my school who are concerned, that this new 11th grade test is culturally biased against African-American youngsters, and the majority of the youngsters in the special needs districts are that.

I believe the test is a violation of equal protection, the doctrine in the New Jersey Constitution, in that it may have a discriminary effect on the graduation opportunities for these special needs district students. I think there is a vital need to study anew the whole issue. You said you had a problem with the early warning, and the 11th grade test, as most people view it, has raised the standards, but it is too steep. It should have been more gradual and should have gone on a yearly basis. From the information I have thus far, that is not what happened. It was decided that they did want to raise the standards -- and we are all concerned about the standards being raised -- but how it is being done, and what will be the effect on the youngsters, is a vital issue.

The commercial tests on the market-- I just don't see why we are duplicating efforts and using additional money, millions, to have a test out of the Department of Education. I am concerned that, you know, so many youngsters, not only for reasons--

A class system, I think, is developing -- is being created more and more in this country. Youngsters in the urban areas, regardless of their racial background, are falling through the cracks. We know many people who are considered low or middle class who are working their way up. Those youngsters live in the city. Their parents have worked hard. They have not been subject to those tests, and then as we go along we see people who are in higher positions who have not been subject to these tests, tests that have been as stringent. Now the youngsters, many of them, have told me that as college graduates, when they graduate from college, and if they are fortunate enough in today's economy to get a job, that later on they find that their superiors do not have as good a background as they do. So therefore, there is a feeling and a perception of unfairness with the testing here.

We are certainly concerned that youngsters have a higher self-esteem. Self-esteem-- We know there is a relationship between self-esteem and academic achievement, because when people have higher self-esteem, it is because they feel as though they have talents, they have achieved something. That is a concern that is very valid. We have been concerned about LA and urban centers throughout this country. I have spent my career in the urban areas, watching youngsters who were very talented who were scared to death of tests, such as the SAT; very bright youngsters, but they may not have been able to score on that type of test to get into a certain type of college.

So, I am concerned about this whole issue, particularly the 11th grade test. The teachers-- And I have looked at it, because as a counselor I am a part of the administration of these tests. Also, the teachers, particularly the math teachers, the reading teachers in my school have said that they are concerned. The big question is: What will happen in December of '93? We will have thousands of youngsters if this test does not change. There will be thousands of youngsters who will fail that test. What is going to happen with those youngsters? What is going to happen with their self-esteem? Will we have youngsters who will decide, "What's the use? I may as well drop out."

We are creating a problem that I think can be limited with some concern about this test being either changed or eliminated or grandfathered, whatever is necessary, because the youngsters in the 10th grade will take-- The youngsters in the year before have taken a much easier test than the ones who will take the 11th grade test. The ones who only had to take the 9th grade test, they have had, from 9th grade through 12th grade, many opportunities to pass that test. Now, their friend may just happen to be, coincidentally, one year, or one grade behind them. They do not have the same opportunities, which goes back to the equal protection. They had much less time.

Even my students are talking about this. The youngsters are intelligent. They may not always test as high as some, but they are smart and they know. They feel as though there is institutionalized racism against them, starting with when they come to school. Certainly now when they are almost out of school, when they have gotten to 11th grade, they feel as though, "My goodness, here it comes again. I have done the best I can, and here at 11th grade when my family -- when we are all looking forward to graduating from high school, I may not make it."

I have administered the 9th grade test over the years, and I have noticed that in my school district the majority -- I can't give you the statistics, but I would say it is in the high 90s, 98 percent or higher -- have eventually passed that test. They felt good about it, because I have gone through it with the youngsters who have failed. Then we had to make sure they had the remediation. I have only had one or two youngsters after a certain period of time and help, who have not finally passed that test. Parents and many youngsters now believe, "Is this test designed to keep a certain group of people down?"

ASSEMBLYMAN ROCCO: Priscilla, your concerns are well-taken. Basically you are talking about what happens to those students who do not pass. We know at this time that they will not receive a normal high school diploma. Will there be an alternate route for them of some type, an alternate route in the sense of a different direction and additional years of schooling to pass? What other alternatives are there?

I think we are very fortunate to have had this hearing today, because I think the Department needs to really prepare, as does the State Board of Ed, for the possibility of students dropping out who do not pass this exam, which would increase the dropout rate in the State. Looking back, as compared to other states, we are so different in New Jersey that we have to

have a totally different test from every other state in the nation. I mean, we are really different. We can't live by the norm tests that are out there. Also, we have really cut off -- potentially cut off -- the ability of many students to move ahead in the future, if they do not pass this exam.

I can tell you, without hesitation-- You know, I know many young people who I grew up with, including myself possibly, who in high school really were not oriented in that direction. But to cut off the ability of an individual to move into the field of higher ed totally-- I think we have to look at that carefully. We have to be ready to look at alternatives, and this Department really has to prepare, because if this 11th grade test -- the New Jersey test-- I not only heard from Priscilla, although she is the one who opened my eyes to it, but from others, that you can potentially have a number of people not making it on this test, as opposed to the national norm tests that are out there that every other state takes. We can relate one state to another. We are going to get better, more precise data that if we take the New Jersey test, we are hindering New Jersey -- New Jersey students and the image of the State.

I won't go into it any more at this time, but I think we will come back to that. I know Joe has some questions.

ASSEMBLYMAN DORIA: Thank you very much, Mr. Chairman. I want to thank John for allowing me to sit in on this hearing and have some opportunity to make a few points.

As John has said -- as Assemblyman Rocco has said -- myself and he have had a number of questions for many years, whether it be a Republican administration or a Democratic administration responsible for the Department of Education, concerning the whole concept of developing a statewide test.

Commissioner, I have a few questions that relate specifically to the process by which the testing has developed here in this State. The first question is: Why-- Obviously,

you were not the Commissioner when this began, that was two Commissioners back, but why was it determined that it was necessary for New Jersey to develop its own test, rather than work with nationally norm tests that could be adapted to the needs of New Jersey?

COMMISSIONER FITZGERALD: Actually, there are 14 states right now, and the number is increasing, which are developing their own graduation tests and moving away from what you would call commercially norm tests, because they are interested in creating a high school curriculum that has been developed inside the state with their own teachers that the test will measure.

ASSEMBLYMAN DORIA: What you are basically saying then, is that it would not be possible to use one of the nationally norm commercially created tests to tie into what would be specifically the curriculum developed within the high schools of New Jersey?

COMMISSIONER FITZGERALD: If you were going to use, like the CAT, or something, that would become our curriculum.

ASSEMBLYMAN DORIA: Is it possible that those companies would be willing to adapt their tests to meet specific needs within New Jersey? As I understand it, the companies are beginning to move in that direction.

COMMISSIONER FITZGERALD: Sure, sure. Assemblyman, actually CTB, which is one of our major test contractors, is one of those contractors which does exactly that for states.

ASSEMBLYMAN DORIA: But the question then again is-- What we are doing here-- We seem to be -- not to use a hackneyed expression -- but we seem to be trying to recreate, you know, an existing test, tying it to our own specific needs, which seems to me, number one, to create problems as it relates to the norming and the question of validity, and also, at the same time, seems to be costing more money that would necessarily have to be spent.

So what we are doing is, we are having problems with the tests, while at the same time we are spending money to create the problems with the test.

COMMISSIONER FITZGERALD: Right. Well, I would say, not having been here, that New Jersey has taken a real leadership role nationally in assessment. Meeting with some people like myself over the last couple of months, I recognize that New Jersey is perceived as a leader in assessment. We are, not just in New Jersey but nationally as a professional, in the midst of an assessment revolution. Tests are definitely being changed. They are moving to performance assessments, all kinds of multiple measures away from a single norm reference standardized test. That is true. New Jersey did take an early look at -- made an early commitment to moving in that direction.

ASSEMBLYMAN DORIA: Then why do you see New Jersey having so many problems? I mean, was it the introduction of the open-ended questions, number one; the lack of validation of those questions that created this type of a situation where we have, you know, 5000 students who have been misplaced into remediation where it was not necessary?

Are you saying then that either the company that did the validation, or the company that is working -- CTB, McGraw-Hill -- if that is the question, the company that was involved that made the mistake-- Did the Department make the mistake? Was it a sharing of problems? Why suddenly do we have this difficulty? Was it because we went to open-ended questions, even though that is what is happening nationally, and movement toward open-ended questions in most standardized tests? What was the reason, do you think? Obviously it is an opinion.

COMMISSIONER FITZGERALD: There were several. Actually, a lot of what you are suggesting, they are all true. There were some judgments made in the Department. There were some judgments made by CTB, McGraw-Hill which I think raised real questions about the equating of that '91-'92 test.

ASSEMBLYMAN DORIA: Could you share with us some of those judgments that were made and why they may have been made?

COMMISSIONER FITZGERALD: Do you want to take this one?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes. One is, I think although we tried to do the right thing -- and I say "we" looking back a couple of years -- we tried to do the right thing too quickly. The time lines for the development of the first year of the EWT were too short. Resources is another issue. There were not enough of them to do field testing in advance. Field testing was done as part of the implementation process through item analysis, rather than in advance. That contributed.

Open-ended questions-- That is the way to go, but to include them in this very precise equating process for setting the cutoff score was not a good decision. They should have been on the test even in the scoring, but not as part of the technical equating process.

ASSEMBLYMAN ROCCO: Why was there the pressure for the time line? That is the first question I would like to ask.

ASSISTANT COMMISSIONER KLAGHOLZ: Well, I don't know. The time line was set in the law itself. That was a couple of years ago. That is, in itself, an answer; that the time line was one year. I think what it was, was that --

ASSEMBLYMAN ROCCO: But the law never indicated that the State had to develop the test. It never said to develop a test by the State. It just said, "a test."

ASSISTANT COMMISSIONER KLAGHOLZ: That is correct.

ASSEMBLYMAN DORIA: Because we voted for the law at that time, with the understanding that it would be given consideration; from the discussions we had with the Commissioner at that time, Commissioner Cooperman, that it would be considered that they use existing standardized tests. The determination was made by the Department after the passage of the legislation that we could create our own tests, as I understand it. Am I correct? That would be the question.

But to go back to the question-- Okay, if the time line was within the test-- Again, the question of why we had to do a separate test is one that was a determination made at the Department level, not at the legislative level.

COMMISSIONER FITZGERALD: Correct, yes.

ASSISTANT COMMISSIONER KLAGHOLZ: That's right.

ASSEMBLYMAN DORIA: Do you know why that determination was made?

COMMISSIONER FITZGERALD: It was related to the core proficiencies.

ASSISTANT COMMISSIONER KLAGHOLZ: Yes. It related to the performance of students on international comparisons, on national tests, SATs, and it was an attempt to improve the performance of New Jersey's students by developing a coherent system in which there would be core proficiencies; there would be a set of curriculum standards that looked beyond the status quo; that would be what we hoped our students would learn in order to succeed in the year 2000 and beyond. And then to develop a testing system that was specifically tailored to measure whether or not students were accomplishing those specific goals -- those specific educational goals -- rather than to draw on a stock commercial test that really is a sampling of what textbooks used in various states offer.

ASSEMBLYMAN DORIA: Was there any attempt, Leo, do you know, to talk to the various commercial test makers about whether or not they would adapt something like the Metropolitans or the CATs, to make those changes necessary to tie it into the curriculum in New Jersey? Did anybody make any attempt to discuss that with these companies?

ASSISTANT COMMISSIONER KLAGHOLZ: Assemblyman, I don't know if those discussions took place.

ASSEMBLYMAN DORIA: Do we know if those companies are interested in doing that at the present time?

ASSISTANT COMMISSIONER KLAGHOLZ: I discussed that with CTB. Essentially their answer was that on the various commercial tests they have available, while they might be able to make some relatively minor accommodations, the more we would ask them to tailor it to our own curriculum standards, the more it would become exactly what we have now, which is a tailored, customized test, and the cost would be--

ASSEMBLYMAN DORIA: But they would be going at it in a different way. I would think that it would be the difference between inductive and deductive. Here we started with our own test and built it up. There it would be taking their test and changing it around. I would think it would be cheaper and maybe more efficient, and probably a little bit more valid.

ASSISTANT COMMISSIONER KLAGHOLZ: Well, their actual answer was that it would become so much a deviation from what they have, that it would be creating a new test, in effect.

ASSEMBLYMAN ROCCO: May I just pick up, Joe, on your question? Isn't it illogical to state that a New Jersey test is going to test us against the international community better than a national norm test that has already been out there?

ASSISTANT COMMISSIONER KLAGHOLZ: No. I think the point was that using those as an indicator of, "Are we doing well or not well in general?" that there was a concentrated effort to improve New Jersey's curriculum and to reassess whether--

ASSEMBLYMAN ROCCO: This is me, Leo; this is me. You know, I know the system. The standardized tests that are out there have a much higher correlation, and would have a much higher correlation to any test that was held in the international community, because they would generally be based upon the standardized tests that are out there. We have a totally unknown test that we are trying to put together, which may or may not have validity or reliability. We don't know that. When we started out, the test was already out there, and

would have a much higher correlation in the fact that it existed and probably was utilized in the past before New Jersey ever had a test. So that is a totally inaccurate statement.

ASSEMBLYMAN CHARLES: May I just elaborate on your point, Mr. Chairman?

ASSEMBLYMAN ROCCO: It's Joe's point, but go ahead, Joe.

ASSEMBLYMAN CHARLES: We are talking about standardized national tests, and we are also concerned about the hurry-up pace at which we were moving away from a national test to a State of New Jersey specific test that had the potential for disadvantaging a large segment of our public school population.

Given those issues, I guess the question is: Was -- and if not, why not -- some consideration given to maintaining the national test, and at the same time moving toward what might have been a New Jersey test, without having that New Jersey test be the standard for graduation? It seems to me that underlying all of this was a determination, a judgment by people within the Department of Education, that it was willing to sacrifice one whole generation of students, perhaps, just so that New Jersey could move forward in a direction they thought was good in terms of educating the student body.

I think everybody agrees that we should move forward with higher standards. The question is, how? One obvious concern is that no segment, or no unnecessary number of young people get disadvantaged by that. It just seems to me that a good way of approaching it would have been to maintain the same systems, and for ourselves develop another test to be used as we want to use it, but at the same time avoiding the stigma that is going to be associated with people getting no degree as a result of tests that are happening here.

See, underlying this is something that concerns me. I don't know how the Department could have made a judgment that

it was, either expressly or inferentially, willing to sacrifice these young people Priscilla has been talking about here. I just don't know how they could do that. We are all here as legislators concerned about what is going to happen. It is a fairness, or an unfairness question.

COMMISSIONER FITZGERALD: I really can't imagine that my predecessors intentionally made a decision to sacrifice a proportion of New Jersey's students.

ASSEMBLYMAN CHARLES: Well, they didn't think about it then. I mean, it seems to me that you're thinking about-- I guess the question should be-- This question should be asked: What was the best guess of those who instituted this system as to the number of people who were going to pass the HSPT? I mean, if they decided that 90 percent of the people were going to pass, well then that is one way of evaluating them. If they could reasonably predict, based upon performances in other tests, that 50 percent of the students, or more, in certain segments were going to fail the test, well then they decided that they didn't care about that part of the population.

COMMISSIONER FITZGERALD: I believe initially a large proportion of students failed the 9th grade HSPT, and over succeeding years that number dropped significantly. I would guess that when we begin the HSPT-11 there will be more students failing it the first year than will fail it the second, third, or fourth year. But frankly I think there was a real commitment not to disenfranchise, but to make sure -- make a greater effort to make sure that students in the inner cities are getting the kind of education that will make them competitive.

ASSEMBLYMAN CHARLES: See, everybody wants that, but it is a question of while you are doing that, don't hurt any more than are already hurt. Everybody wants that. I am one for the highest possible standards. I want the standards to be as high as we can have them for everybody, with nobody

skating. But in the meantime, until we get to that point, let's do it in a way that doesn't bring any more hurt and pain and disadvantage than already exists. I think that is what everybody is talking about here.

ASSEMBLYMAN DORIA: I think, let's get to the genesis of the question: I think the genesis here has to be, what was the reason that the Department made the decision to create its own test and not do a period of phase-in or change, as Assemblyman Charles said? The legislation says only this about the test: "The test shall measure those basic skills all students must possess to function politically, economically, and socially in a democratic society." That is a very broad and very general statement.

The question then becomes one that is now not a determination of legislation, but rather a determination of the departmental regulation and decision of the State Board of Education. The question is: How much imput did the State Board of Education have in the final determination, and did they? Or was this an administrative bureaucratic decision that we should move immediately to create our own test, doing exactly what Joe just pointed out, which I think is a reasonable question, which ties into Assemblywoman Anderson's question, and that is, could we have phased in, over time, a very stringent test, which ties in with what Assemblyman Wolfe and Assemblyman Rocco said; all of them saying the same thing in a different way -- and Assemblyman Nickles?

The question is: What was the necessity of pushing through this test so quickly that created these problems, and why did we not take some time to develop a test that would be appropriate and phase in using, first, standardized test adaptations, and then move towards the creation of a unique test for New Jersey?

That you cannot answer, Commissioner, because you were not there, but I think the Department-- That is a question

that I think all of us would like to have answered by the Department, a historical question, of why Commissioner Cooperman, Commissioner Ellis, and now moving to the present, made that determination, given the fact that the legislation was so general in its mandate, and given the fact that all of us feel that excellence is the priority. We all want excellence, but we want excellence that is developed over time. To demand that suddenly you move and do it immediately is not realistic, and then to demand that a test be created which has to meet these needs, which suddenly is created uniquely in New Jersey--

I know it is great to be the first to do things. We all like that, and we all like to say that we are the leader in something. But the question is-- If we are the leader and it doesn't get done in a proper manner, I think all of us have a question. I don't know if there is an answer to that, but I think that is an important question for all of us to ask, and I think it is an important question for the Department to ask.

Maybe, Mr. Chairman, with your concurrence, that might be something that could come back to us in writing at some point.

ASSEMBLYMAN ROCCO: I think that is very critical. From the questions that are coming out of the Committee today, I think this is probably going to require additional time with some responses, as well.

David, before I go to you, Rudy has not had his time, and Joe has not finished.

ASSEMBLYMAN DORIA: Let me just ask one final question. This is the final question I have to ask, Commissioner.

ASSEMBLYMAN ROCCO: Then we have Bob, and then we have David.

ASSEMBLYMAN DORIA: Why this year did we go from a band to a specific cutoff score? Again, the same question of

why suddenly do we have to go full steam ahead? If there are questions with the validation, why do we have to go from a band that allowed for a determination of where the problems existed, which creates some flexibility which is necessary, to suddenly creating one specific cutoff score? What is the answer to that?

COMMISSIONER FITZGERALD: We have the answer. It relates to both the monitoring and the fact that it is a graduation test. Do you want to detail that?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes. Well, first of all, the EWT. We went from bands to a cutoff score. The experts we brought in said that we should not have done that, so we are going back to bands.

ASSEMBLYMAN DORIA: Yes, it makes more sense when you are trying to come up with, you know, remediation. Why suddenly just one number? It is like, you know, suddenly creating an arbitrary date for when we have to put in a tax form. It's nice; there is sense to that, but why, when you are dealing with a case of remediation -- not a case of graduation-- I can understand the 11th grade test, but why the question of the 8th grade test, which is basically a test to prepare? What was the reason, Leo? I mean, why do we do it? I understand why we are not going to do it, but why do we do it?

ASSISTANT COMMISSIONER KLAGHOLZ: That's a harder question to answer. I am trying to think back on our investigation into this thing. I think it was a well-intended attempt to match it up to what would happen ultimately on the HSPT, but again my own view, and the view of the people we brought in, was that it was a misguided attempt.

ASSEMBLYMAN DORIA: Okay. I just want to thank you, Commissioner, and Leo, for your forbearance. And I want to thank the Chairman again for allowing me to join in. I think this is a serious question; one that all of us have to be concerned about, because it doesn't only impact upon the students today, but upon all students. I think we have to be

realistic enough to deal with the fact that sometimes we have to take time to do things, rather than rush to try to make things look good. I think that is a problem that always exists among all of us, including myself.

ASSEMBLYMAN ROCCO: Also, Joe, I think we have to be concerned about throwing good money after bad. Basically, you know, we are talking about $2 million this year, and I assume it will be somewhat more by the time we finish. Then we have years to follow in which we are going to put millions and millions in trying to be the New Jersey test, when there are companies out there that have already done that.

Before I go to Rudy, who has not presented any questions as yet, next year you are going to a 4th grade test as well?

ASSISTANT COMMISSIONER KLAGHOLZ: I'm sorry?

ASSEMBLYMAN ROCCO: There is going to be a 4th grade test next year?

ASSISTANT COMMISSIONER KLAGHOLZ: No, there won't be. There isn't money to do that.

ASSEMBLYMAN ROCCO: There will not be a 4th grade test. There will just be the early warning--

ASSISTANT COMMISSIONER KLAGHOLZ: The district selected test at the 4th grade level, which is required.

ASSEMBLYMAN ROCCO: Each district presently gives-- You know, I taught 4th grade, and we gave the tests.

ASSISTANT COMMISSIONER KLAGHOLZ: Right, exactly.

ASSEMBLYMAN ROCCO: Can you guys live with that one? If you can live with it there, why can't you live with it in these other areas? But I will get back to that.

Rudy?

ASSEMBLYMAN GARCIA: Thank you for coming here today.

My question just ties in with a lot of remarks that were made here today. It regards the HSPT test and LEP students -- the limited English proficiency students. Many of

these students-- A lot of them live in my community, the community I represent, the 33rd Legislative District. Many of them are new arrivals to this country.

I believe that some of the problems that the students across New Jersey have faced dealing with this whole testing situation are compounded for these students, who, by definition, are limited in their understanding, application, and use of the English language.

My question is simply: Why do we put these students in a position to fail? I mean, obviously they are going to fail this test. It goes back to the whole concept of self-esteem. That is probably the greatest example that this test was just pushed forward upon this whole class. Being a student myself not too long ago, going into a test where you know you are going to fail really affects you. It affects you emotionally; it affects you psychologically. These students are being made to take this test. Many of them can't even understand the directions on the test, let alone the contents or the questions asked. I just want to see basically why were these students put in this position, and what role do you see these tests playing for these LEP students?

COMMISSIONER FITZGERALD: Well, Assemblyman, I have to tell you that I agree with you. I met with the Bilingual Advisory Committee several weeks ago, and we now have a committee that is coming back to me with recommendations on April 25 to change it.

ASSEMBLYMAN GARCIA: I saw some of those recommendations. Are they going to be done away with for these students?

COMMISSIONER FITZGERALD: We'll have to review them. You can make a very strong argument. It is certainly reasonable to think that they shouldn't be taking a test they can't even read.

ASSEMBLYMAN DORIA: There is some logic to that.

ASSEMBLYMAN ROCCO: A little logic.

ASSEMBLYMAN GARCIA: If it is so obvious to us, why-- (everyone on the Committee speaking at once here; indiscernible)

ASSEMBLYMAN ROCCO: Remember the old saying, "We're from Trenton. We're here to help you" (laughter)

COMMISSIONER FITZGERALD: I don't know, Assemblyman.

ASSEMBLYMAN GARCIA: That just goes to the whole thing about self-esteem and how these students were really not thought of.

ASSEMBLYMAN ROCCO: Thank you, Joe Doria, for being with us.

Rudy, you were finished?

ASSEMBLYMAN GARCIA: Yes, I just wanted to touch on that.

ASSEMBLYMAN ROCCO: Vice Chairman Martin?

ASSEMBLYMAN MARTIN: Let me just start by-- One of the things that occured to me-- One could also draw another interpretation of the reason why New Jersey has gone into having its own tests. The fact that it's a unique test that cannot be compared against any other state or against any other norm, also lends itself to not being able to determine what its real value is -- how students compare with students outside the State. Also, drawing whatever cutoffs -- to use your phrase -- as to who should pass or not, also seems to me somewhat arbitrary in any given year on a test that no one else is familiar with.

I'm as concerned that it could be a test that could present the results it wants, since no one else is using it and there is no real measurement against anybody else, as it is that it could be directed to be a test that nobody could pass. Is that a possibility with our own unique test?

COMMISSIONER FITZGERALD: Well, we're obviously using other kinds of measurements in high schools to really assess how well our kids are doing. If that was the only thing we

were using in New Jersey, I would say, "Yeah, you have a good argument," but kids are taking a lot of other things that do have national norms to them.

We're requiring 10th, 11th -- 10th graders, 12th graders -- to take a norm reference test, so we have data on nationally normed tests. They also certainly take the SATs, they're taking advance placement tests, and all of those things have national norms to them.

ASSEMBLYMAN MARTIN: I guess that goes to what John said. If we're going to rely on those to be able to determine whether our tests make sense, why do we need our separate tests?

COMMISSIONER FITZGERALD: Well, the issue is, it is a graduation test, and legally we would have to make sure-- I know you understand that the content of the test is our curriculum, or you can't use it to confirm or verify that they can be graduated from our high schools.

I mean, this is supposed to determine or say, "They know what they were taught."

ASSEMBLYMAN MARTIN: I understand that, but as far as the passing levels on tests like that, to be able to set it and say, "This is the 11th grade of our own internal system, as to what an 11th grade student should know or not know," seems to me, without some kind of check on it, makes it open to somewhat arbitrary decision making.

COMMISSIONER FITZGERALD: Right. You wouldn't really use an arbitrary system.

ASSEMBLYMAN ROCCO: Suffice to say, Commissioner -- if I may, Bob -- that the nationally normed test also feed into the curriculum. They're the types of things that are taught through each state at each level. You know, they are certainly part of the curriculum and always have been.

I think what we are really doing here is we're developing a test that drives a curriculum instead of having the curriculum drive the test, in this process.

ASSEMBLYMAN CHARLES: Mr. Chairman, why do they have to be the same, even? I mean, it occurs to me that if we have a test -- a national test with norms, basic, with minimum skills, let's say, that's what somebody develops nationally, so we ought to be able to do that. Our kids ought to be able to pass a minimum test.

Now, if we want to teach them more, that's fine. We ought to go on and develop curriculum that teaches them more, and makes them better than everybody else. We can teach them that, and know that we're teaching them that. Give them the same national test, they'll pass that with flying colors.

ASSEMBLYMAN ROCCO: And use that as the marker.

ASSEMBLYMAN CHARLES: Right, and we'll still be doing the job. We can figure out our own way of determining whether or not our curriculum-- Well, that would be an indication of whether or not our curriculum is really doing a better job, because our kids would score higher on this normative test.

I don't see that the curriculum-- The test is supposed to measure what you taught, fine. But that doesn't mean that we have to have this test as the determinative of whatever it is of measuring our kids, especially when that same test score is going to be compared to what other people measure in their scores, or may not be measurable, as the points that Bob just pointed out.

ASSEMBLYMAN ROCCO: Right.

ASSEMBLYMAN CHARLES: You know, too much curriculum tests-- I don't see the compulsory--

ASSEMBLYMAN ROCCO: Joe, most nationally normed tests -- and I know Bob wants to get back here -- but most nationally normed tests have your reading skills, your English, your math, your math computation.

ASSEMBLYMAN CHARLES: But, those are minimum, sure. We know that everybody should know a certain amount of ABCs and reading and all of that, sure.

ASSEMBLYMAN MARTIN: I wanted to ask about the ABCs. As I understand the test, and correct me if I'm wrong, there was an attempt by Governor Kean to stress the development of various core proficiencies. I don't think we've followed that model, but this test, unless I'm wrong, stresses three areas, which are reading, writing, and mathematical skills. Is that right? Is that what it concentrates on?

COMMISSIONER FITZGERALD: That's correct.

ASSEMBLYMAN MARTIN: So when we talked before about that it measured the broad curriculum, it measured the basic--

ASSEMBLYMAN ROCCO: It doesn't measure the broad curriculum.

ASSEMBLYMAN MARTIN: --three "Rs" that Joe alluded to.

COMMISSIONER FITZGERALD: That's correct. That's correct. It measures differently at another level than you can currently find available in commercially prepared tests, especially in the area of mathematics. Mathematics has really undergone a revolution in the last decade using the new math standards, and those definitely are now in our curriculum and on our 11th grade test.

ASSEMBLYMAN MARTIN: Was there any attempt to use the basic skills that are used in the New Jersey colleges, which also tests those, as I understand, pretty much the same as-- I have a little bit of familiarity because my wife does teach at one of the colleges in the English Department, and does go through that assessment. Are there open-ended English questions, at least for writing? Was there a test to try and correlate the two, as far as this test goes?

COMMISSIONER FITZGERALD: No.

ASSISTANT COMMISSIONER KLAGHOLZ: No, there was not.

ASSEMBLYMAN MARTIN: Because theoretically, then, you could have a student who passed this, but go into Ocean County Community College, or Montclair State College, and be determined-- Because unless the test correlates 100 percent,

they could be identified as still having problems as far as their basic skills.

COMMISSIONER FITZGERALD: Yes. One of the recommendations from the Governor's Management Review Commission Operational Audit of the Department was that perhaps one of these tests ought to be eliminated. The 11th grade test, now that it's at the 11th grade, is a much more difficult test than the 9th grade test, and probably the two tests are much closer than they were.

ASSEMBLYMAN MARTIN: Is that just a problem of -- in your view -- of jurisdictional, between the Department of Higher Education and the Department of--

COMMISSIONER FITZGERALD: Well, I've had a preliminary discussion with the Chancellor about that. I don't think it would be jurisdictional if we made the decision.

ASSEMBLYMAN MARTIN: But, hypothetically, if there were one department, and you were calling the shots, you would move it in the direction of only having one and the same test?

COMMISSIONER FITZGERALD: It sounds very reasonable to me.

ASSEMBLYMAN ROCCO: I know David is at a county college. He wants to respond to that, too, as well, Bob.

ASSEMBLYMAN WOLFE: I had to wait to ask my questions until last, and that's what I wanted to ask 10 minutes ago, exactly what Assemblyman Martin asked. Now, you indicated the possibility of eliminating either the 8th or the 11th grade tests, or are you talking about combining with--

COMMISSIONER FITZGERALD: The basic skills --- the college basic skills.

ASSEMBLYMAN WOLFE: Okay. That was my point, because it does seem redundant.

COMMISSIONER FITZGERALD: Yes, it does.

ASSEMBLYMAN WOLFE: Because it is a test of minimal basic skills.

COMMISSIONER FITZGERALD: If they pass the 11th grade test, they really ought to be able to pass that minimum basic skills test.

ASSEMBLYMAN WOLFE: Right. But the point is, why give two tests? Why not use one for either, or for both?

ASSEMBLYMAN ROCCO: I think we'll pursue that more. Bob wants to continue, though.

ASSEMBLYMAN WOLFE: Good, thank you. I'd like to find out about that. How can we find out about that?

ASSEMBLYMAN MARTIN: The feasibility of that--

ASSEMBLYMAN ROCCO: We can have the Commissioner and the Chancellor discuss it. I'm sure that's a possibility.

ASSEMBLYMAN WOLFE: Would you discuss that over lunch? (laughter) Thank you.

ASSEMBLYMAN ROCCO: Bob still has the floor.

ASSEMBLYMAN MARTIN: This, as you call it, the SRA mechanism, which is the alternative testing procedure. What does SRA stand for?

COMMISSIONER FITZGERALD: Student review assessment.

ASSEMBLYMAN MARTIN: Reassessment?

COMMISSIONER FITZGERALD: Student review.

ASSEMBLYMAN MARTIN: Has there been some attempt to create validity of that procedure which is going to take place next year, and is it going to arguably give thousands of students-- Or going to be judging whether they receive a diploma or not?

COMMISSIONER FITZGERALD: We're in the process of that. Yes. That's been a part of the creation of the new task, just like when we created the 9th grade test we had to create simultaneously an SRA for that test, so we have been in the process of creating the SRA review for the HSPT-11. Yes.

ASSEMBLYMAN MARTIN: Because, it seems--

ASSEMBLYMAN ROCCO: That will take into account test phobic children as well?

COMMISSIONER FITZGERALD: It has. Yes.

ASSEMBLYMAN MARTIN: Isn't that the basic argument as to why they would pass on that: because they don't test well?

COMMISSIONER FITZGERALD: That could be one reason, but as some other people have suggested, sometimes these tests are gender and racially biased.

ASSEMBLYMAN MARTIN: Well, I guess my concern is, it has a potential of being highly subjective.

COMMISSIONER FITZGERALD: Which, the SRA?

ASSEMBLYMAN MARTIN: The SRA, because that's the sort of the-- There are going to be several-- Conceptually, is this going to be several persons where the student who has for whatever reason not passed the standardized test, going to go before them, as a Committee? How do you envision this? I mean, it's only a year away.

COMMISSIONER FITZGERALD: Well, the SRA procedure is in place because we've been using it with the 9th grade test. As a superintendent, I signed them when they finally came to me, and they were a very thick document. There needs to be a lot of documentation through the guidance department. Isn't that right?

ASSEMBLYWOMAN ANDERSON: Yes, and special teachers in each area to work with those students individually.

COMMISSIONER FITZGERALD: Right. It's quite extensive, through tutoring after school. Isn't that right? Study halls, depending upon the high school--

ASSEMBLYMAN NICKLES: If I may, Mr. Chairman. Basically, it's a finding of fact with documentation to support the evidence that a graduation diploma should be issued in lieu of not passing the test.

COMMISSIONER FITZGERALD: That they know it. Yes, that they know the information.

ASSEMBLYMAN MARTIN: Is there going to be any attempt at the 11th grade level to have some person or process outside

of the school district itself to make that determination? Is it going to be all internal?

COMMISSIONER FITZGERALD: Well, I believe the County Offices review the SRA with the districts, before signing off. That has occurred in the past, and that will occur in the future. This is a very rigorous process. Those things haven't happened easily. I know that as a recipient of the direction from the State Department, and we never took it frivolously.

ASSEMBLYMAN MARTIN: I'm not suggesting you did. I'm just concerned about a process that may be overwhelmed next year in making decisions that are going to affect the students.

COMMISSIONER FITZGERALD: Or the next year. That's correct.

ASSEMBLYMAN MARTIN: Because if somehow word gets out that it's an easy process, then it will hurt the rigor of the test. If it's understood that somehow it's biased and students don't think they're going to get a fair shake in coming before it, it's also going to have very damaging repercussions. I would hope that, you know, at least at this stage that that be developed and closely monitored.

I just want to make a comment; it's related to Priscilla's. I think, at least in reflection, the testing procedure should be done, in my view, probably most at the 8th grade level and not at the 11th grade level. And there's a couple of reasons, one of which-- I don't think it's been said before, but since we know that students can leave the school system at 16, and at 18 they become adults and can make their own decisions, one of the problems that you have is when a student is so close to the end and they realize that they may have too much work to do -- and there may be other pressures, especially in some of our social environmental areas that have problems -- that it may be too late.

Whereas, at least if there was a rigorous 8th grade test, then I would think that a student who did well there, and

there was enough time to get that student to be able to catch up after the fact, then I think we'd be in better shape to salvage that person's career and their sort of well being, their mental adjustment. Maybe we're misplacing instead of the test, maybe we're misplacing our emphasis, because the time to catch these people and work with them, it seems to me, is earlier, not later.

COMMISSIONER FITZGERALD: Well, we do have-- That rigorous test you're talking about is the 8th grade EWT which is a rigorous test.

ASSEMBLYMAN MARTIN: I mean, I'm not saying-- Maybe at that point we should say who should go on to high school or something, maybe build other incentives or something like that, but my point is not to take somebody-- It always bothered me, with a Ph.D. -- I don't have one -- but when you looked at the process, somebody takes, like-- There's a potential right at the end to snatch it away. I don't want to see that developed through giving out signals to our students. It should be very positive and they should have an opportunity. I think we may be placing our emphasis too much at the end. Thank you.

ASSEMBLYMAN ROCCO: David, did you finish your question?

ASSEMBLYMAN WOLFE: Thank you. Yes.

ASSEMBLYMAN ROCCO: Priscilla?

ASSEMBLYWOMAN ANDERSON: One of the problems, Commissioner, that we noticed with the new 11th grade test, is that it is geared towards youngsters -- it appears to be geared more towards youngsters who are going to continue post secondary education, and that's good to prepare those who are going onto higher education, as Assemblyman Wolfe says. But many times, particularly in the special needs districts, perhaps the majority of youngsters do not have that interest that they may want to go into other kinds of careers and skills, and so on. So that is the problem, in that those who

spoke to me were thinking that it was like a college level test, so the youngsters who are preparing in that curriculum -- academic, college prep curriculum -- would be more prepared for that test. But there are other youngsters who are in both vocational, general, clerical, and business, that just want to go to work when they graduate. They're the ones where the concern is very high.

COMMISSIONER FITZGERALD: Right. And there are 45 percent of kids in New Jersey who don't go on to college.

ASSEMBLYWOMAN ANDERSON: Exactly.

COMMISSIONER FITZGERALD: I know that. Of course it's a little difficult anymore to get a job that isn't requiring a lot of these skills. I mean, that's part of the problem, and to make sure that kids have got what they need to get a decent job and not something that's minimum wage, I think, is really the motivation here.

ASSEMBLYWOMAN ANDERSON: I understand that. I think all of us agree on that. We want to see them achieve.

ASSEMBLYMAN ROCCO: Before I go through my list, are there any other Committee members at this time-- What I'd like to do is just run through a list of specifics and see where we are at this time, if that's okay, Commissioner? Then I think that should be sufficient.

Would you say that there's a higher correlation between the nationally norm tests that are out there versus the New Jersey tests with the true measurement for students?

COMMISSIONER FITZGERALD: A higher correlation-- I don't understand.

ASSEMBLYMAN ROCCO: Is there a higher correlation? In other words, do you think that the New Jersey test correlates as well as the nationally norm test?

COMMISSIONER FITZGERALD: With what?

ASSEMBLYMAN ROCCO: With what the test is supposed to measure?

COMMISSIONER FITZGERALD: I think it is just as valid as the nationally norm test, yes.

ASSEMBLYMAN ROCCO: You truly believe that?

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN ROCCO: Are we headed for further possible errors as we have had with 5000 students that were, regardless of what you call it, misplaced, because ultimately the principals or superintendents placed them--

COMMISSIONER FITZGERALD: I understand.

ASSEMBLYMAN ROCCO: Do you think the continued possibility of that occurring simply because of the resources of the State, or the fact that we have State employees that are doing nothing but the testing program, do you think there's that potential still existing out there?

COMMISSIONER FITZGERALD: Well, number one, we're not going to have a cut score. We're not creating a cut score for the '93 test, but we would have a performance band so you could really get yourself in the same situation using a band, rather than a single cut score. I don't want to simply, downplay this.

ASSEMBLYMAN ROCCO: I guess if the new 11th grade HSPT really is taking reading and math, basically, and some open-ended questions, I mean, what is really the difference between the tests that are in the field and the tests that we have in New Jersey, other than the written aspect of it or a few open-ended questions? Is there anything really there that's significant that would be worth subjecting students to error, or as Joe has indicated, to subjecting them to -- and Priscilla, basically, to the fact that these tests are not nearly as reliable or valid as the tests that are in existence?

COMMISSIONER FITZGERALD: Currently, the testing industry is undergoing a major transition. If it were five years from now, I think we could, indeed, find some commercially available tests that used the new standards in mathematics and were using applications that were better tests.

ASSEMBLYMAN ROCCO: Commissioner, are you indicating there are companies out there that really are not testing the new math?

COMMISSIONER FITZGERALD: I don't think they're as far along as our own tests. I don't think we have yet-- I've talked to a lot of these testing companies myself, trying to find a new standardized test for Montclair, and I know what their publishing dates are. And they are, of course, because there's going to be a big market, but right now or when we first created this test, I do not think they were available.

ASSEMBLYMAN ROCCO: Well, I would think and assume, and be willing to bet, basically, that there are tests that are out there that will be out there in a year or so. Are we, in fact, spending too much of our time, money, resources, energy, staff personnel, when this Department needs to be really restructured and put back on the right path? Are we just spending too much of our energies trying to be test makers when we should, in fact, be doing other things?

COMMISSIONER FITZGERALD: It's a really difficult question for me to answer. I've just been here three months. I've inherited a lot of things that already were initiated.

ASSEMBLYMAN ROCCO: Yes, you have.

COMMISSIONER FITZGERALD: I don't think it's a bad test. Two years ago when we were discussing viabilities, I don't really know--

ASSEMBLYMAN ROCCO: I guess that's not my question. My question really is, are we spending too much of our time, energy, resources, money, personnel, facilities, etc., trying to be test makers? Are we out of the realm of our business, which is to be the Department of Education: to oversee the schools in the State of New Jersey, to use our resources that are so badly needed in so many ways? Are we just kind of like just in the wrong ballpark here?

COMMISSIONER FITZGERALD: Well, I understand your question. I have to say the test, at this point, is already created, so we really aren't spending a lot more time in the future because they already exist. Except for two or three other--

ASSEMBLYMAN ROCCO: These tests, I can assure you, will bring a howl throughout the State. You're going to find that we're going to be doing -- and trying to be doing -- so much work in this Department, to do the things that Priscilla and Joe and some others have brought up, in trying to find ways to make this test fit. We're a long way from being over, in my estimation, with this test. So, I go back: Are we just in the wrong ballpark?

COMMISSIONER FITZGERALD: But I just work here. I would prefer keeping the tests. We spent a lot of money. A lot of time and energy has gone into creating something that a lot of people are pretty proud of. They're definitely cutting-edge tests. Every consultant I've shown these things to are very complimentary about what we've created. It isn't three years ago, and I wasn't initially at the table when we were discussing options, but, at this point, I think there's more to be gained by going forward than taking a step backwards.

ASSEMBLYMAN ROCCO: Isn't that part of the pride of authorship? I mean, the State wants to be the leading edge.

COMMISSIONER FITZGERALD: Well, these aren't my tests. I wasn't here, and you know, pride of authorship-- I was out there resisting just like Fred. (laughter)

ASSEMBLYMAN ROCCO: I know that, Commissioner. I know that. I guess I'm just trying to focus on the issue of the fact that, I feel, Joe Doria feels, and I think the others on this Commission feel, that we're just so weak in so many other areas, and we need your help in this State. In so many other areas we need your expertise, and the resources of this Department have been depleted because we have spent so much

time trying to be a test maker. I think that is absolutely the wrong way to move, and continue to move, because I think we're a long way from being finished. I think we're going to come back, and you're going to find that we're going to have to redo everything, because there's going to be an upheaval in this State when that testing plan takes place.

I predict that we are going utilize more resources much more than we should be, when, in fact, if we had nationally norm testing: "This is it. See how you do. See how you measure up across the nation. See how you measure internationally," and we'll use the markers from there. We can use other tests for other purposes, but for the high school graduation, this is the test we're going to use, the nationally norm test that we all know the curricula throughout the State would fit into, because they're the kind of things that we're teaching. Because that's why they make the test: to fit into a national type program.

I, for one, would suggest highly that we're just not doing the job that has to be done, and we're not capable in the future of correcting the errors that have been made. I would leave that on your doorstep. I think that is an absolutely critical issue that the Department has to deal with. You know, should you really be out there trying to help the schools and get these districts some of the personnel they need? The County Offices are depleted. Do they need some direction? They need help in that instead of sitting there trying to make a test to fit something that would make the Department feel good about pride of authorship. I have strong feelings in that regard.

As we move ahead to next year, you know, we're going to depend on some standardized tests that are already out there for certain purposes. I know in speaking with you, in the 11th grade High School Proficiency, now you're talking about a fall/spring -- fall/spring, giving the students four

opportunities to pass. Right? That may or may not be enough to do it. Time will tell, I would assume. Have you determined cutoffs on these tests, yet -- the 11th grade proficiency?

COMMISSIONER FITZGERALD: No.

ASSEMBLYMAN ROCCO: Okay. So that obviously, you're going to have to look at that. How are you going to determine cutoffs?

COMMISSIONER FITZGERALD: There are two or three statistical methods of doing that.

ASSEMBLYMAN ROCCO: Such as?

COMMISSIONER FITZGERALD: Do you mind if I call on my Director of the Assessment Bureau?

ASSEMBLYMAN ROCCO: Certainly.

COMMISSIONER FITZGERALD: Well, let me say this, because we don't have to get too technical here, one involves using the first administration of the test to help set scores, which is what any norm reference test does. It tests and then looks at the sample, which is one reason why we haven't set the test scores yet. We have to cut scores yet.

I would be happy to provide the technical information for you, but they are statistical procedures that every test maker uses to establish a cut score.

ASSEMBLYMAN ROCCO: How many people is the Department going to be using in the next year to work on testing and testing programs?

COMMISSIONER FITZGERALD: The Department is quite diminished. I believe we have six professionals remaining.

ASSEMBLYMAN ROCCO: And those six people are going to kind of attempt to finish up at this stage?

COMMISSIONER FITZGERALD: Well, they have other responsibilities beyond the 11th and 8th grade. That's our entire Bureau of Assessment.

ASSEMBLYMAN ROCCO: At the same time the monitoring is going to really kick in, basically?

COMMISSIONER FITZGERALD: The monitoring begins July 1.

ASSEMBLYMAN ROCCO: You'll have to make a choice between the two. I'm trying to say to you that there are options on your part, too, that we think makes sense.

COMMISSIONER FITZGERALD: We actually have some ongoing obligations beyond those two tests. Of course you know we're requiring right now an assessment on every grade level, and we review at grades 4 and 8. So, the Bureau has a rather major responsibility right now inside the Department.

ASSEMBLYMAN ROCCO: Yes, but you're reviewing based-- But every-- A specific test, or each district has their own Californias, Iowas, or what?

COMMISSIONER FITZGERALD: You have a choice.

ASSEMBLYMAN ROCCO: District choices?

COMMISSIONER FITZGERALD: Uh huh. (affirmative response)

ASSEMBLYMAN ROCCO: Okay.

COMMISSIONER FITZGERALD: Which is what we did with the 4th grade test, as a matter of fact. You asked about that a minute ago. Originally the Department was going to create a 4th grade test.

ASSEMBLYMAN ROCCO: Right.

COMMISSIONER FITZGERALD: We have decided we're moving into the direction of simply letting districts use their own 4th grade test, and report those scores.

ASSEMBLYMAN ROCCO: Right. I think they had a 1st grade test a while back, did they not?

COMMISSIONER FITZGERALD: The State?

ASSEMBLYMAN ROCCO: The State. Wasn't there a first grade test?

COMMISSIONER FITZGERALD: No. At this point, K through 1 and K through 2 is totally local, and scores are not reported to the State.

ASSEMBLYMAN ROCCO: I think there was a State test a way back a few years, and--

COMMISSIONER FITZGERALD: Sometimes for Chapter 1, What happens is, for Chapter 1 you've got to test kids.

ASSEMBLYMAN ROCCO: I guess I can't--

By the way, Priscilla's leaving us, as is Alan, because they're going to the new Commission to work on the funding for the State of New Jersey. So, we wish you well in that regard.

I guess what I'm leading up to is that I think as the new Commissioner, we're trying to provide you with alternatives. We're trying to let you know that we feel -- basically, a number of people on this Committee feel -- that there are ways to use the resources that are different than they have been utilized in the past, and would suggest that they be given consideration in that regard.

ASSEMBLYMAN MARTIN: I just wanted to say something while I have an opportunity. I want to follow up with you directly, if I can at some time, Commissioner, but you mentioned monitoring is occurring this year. I have some legislation in. It's preliminary, but the concept of it is that we would not have statewide monitoring, but rather monitoring would be exempt for districts provided they could show some degree of general proficiency, and the Department would be able to concentrate its resources on those districts it felt needed monitoring because they have particular problems. I would hope I could talk to you about that.

There's a related bill that I've been working on -- it hasn't been introduced yet -- that would take some of the criteria the Department has spent on construction, which I think may be excessive as far as regulations and facilities, and leave more of this to school districts. I mention that because, if you could spend less time on providing regulations that go beyond New Jersey Construction Code and the BOCA Code,

and you could concentrate more on some of the things that John alluded to instead of monitoring, especially for school districts which have a good track record within them, I would hope that we could relieve you of some of those burdens that you have just suggested.

I would like to talk to you about that, and some ideas I have about teachers and professional development. Maybe at some point in the future we can get together and sort of pursue those, but I know that you are overburdened right now. These are just some thoughts that I have that would make the Department be able to focus more on what I think you should be doing, as opposed to being an overseer of all kinds of things, as John suggested, which are maybe just too much and unproductive.

ASSEMBLYMAN ROCCO: Those issues that Bob has brought forth are issues that I'm also working on, as well, and I think Jack Ewing is, so I think there's a lot of interest in the issues that Bob has talked about which could take another total morning, at least.

I want to get to some answers, though. In Higher Ed, the combination of tests -- are you going to initiate, or must we initiate some move to combine the testing; the Higher Ed testing with the Department testing? Or is this something you're looking to--

COMMISSIONER FITZGERALD: As I said, I had a preliminary discussion with the Chancellor, and I want to follow up on it. I'll certainly get back to you.

ASSEMBLYMAN ROCCO: Could you kind of give us some information as you move along in that, please?

COMMISSIONER FITZGERALD: Yes.

ASSEMBLYMAN ROCCO: Test validation we've talked about. Money in the future, I assume, is going to continue to be required to a great degree, and I don't think it's money well spent.

I want to go back to the issue of notification of those students. This one letter is just a sample of a district that's going to, again, be placing this child in remediation unless they get notified this child was misplaced. Do you have any suggestions? Do you think it would be appropriate to notify those districts? Are they going to be notified?

COMMISSIONER FITZGERALD: I'm sorry, I didn't pick up your question.

ASSEMBLYMAN ROCCO: The 5000 students that were misplaced?

COMMISSIONER FITZGERALD: Did you receive a copy of the letter I sent to the superintendents?

ASSEMBLYMAN ROCCO: Yes, I did.

COMMISSIONER FITZGERALD: It said that districts and local boards should be reviewing, individually, the students that are in that band, and make an individual determination as to whether or not they ought to continue.

ASSEMBLYMAN ROCCO: All right, so you're not requiring notification, but you are requiring that they review it, and where placement needs to be adjusted, so adjust.

ASSEMBLYMAN MARTIN: I just have one-- John is suggesting something. He's talked about his own early warning system about the tests next year, but God forbid we have some questions in that 11th grade test next year that are later determined by professionals to be invalid, and affect somebody's-- I mean, there's clearly potentially liability questions. It puts the whole system into question, and if we're only relying on our own internal testing, we don't have much comfort being able to draw on the educational community as a whole. I'm sure you're aware of that. That's one of the things, you know -- how we notify these students. At least it's not the final chapter in their career in New Jersey's public school system, but next year's may well be based upon some question that has some arbitrary logic to it, as far as right or wrong.

COMMISSIONER FITZGERALD: I know that.

ASSEMBLYMAN ROCCO: As well, there's no State money for remediation, I understand, other than the at-risk money?

COMMISSIONER FITZGERALD: Remediation? There is Chapter 1 money and at-risk money.

ASSEMBLYMAN ROCCO: Is the new New Jersey test the new curriculum for the State of New Jersey?

COMMISSIONER FITZGERALD: I'm sorry?

ASSEMBLYMAN ROCCO: The new New Jersey test. I mean the 11th grade proficiency. Will this become the curriculum for the State of New Jersey?

COMMISSIONER FITZGERALD: The core proficiencies were written first for the high school level. High school teachers developed those over the last four or five years, which are being measured on the test.

ASSEMBLYMAN ROCCO: So each district now will receive these core proficiencies? They have the core proficiencies?

COMMISSIONER FITZGERALD: They're already out. Over the last couple of years they've been out.

ASSEMBLYMAN ROCCO: They will then utilize these core proficiencies in terms of their curriculum?

COMMISSIONER FITZGERALD: They have been.

ASSEMBLYMAN ROCCO: They have been, and will continue to utilize them in terms of preparing for this test.

What type of preparation are we having for the possibility of the increased dropout rate, or preparing for those students that do not make it on the 11th grade proficiency?

COMMISSIONER FITZGERALD: One of the things I've started doing is meeting directly with high school principals, which I think is a very good initiative, to explore some of these issues. I say that because my own high school principal really didn't have any constituent group to work through some of these problems with. The county superintendents meet with

superintendents, not with principals, and I think a lot of responsibility is falling on these high school principals to pull this off. They definitely need a lot more opportunities to receive guidance and direction from the State Department.

ASSEMBLYMAN ROCCO: Okay. The Department is really going to have to kind of be the overseer, obviously, to make sure that preparation is occurring.

COMMISSIONER FITZGERALD: Right.

ASSEMBLYMAN ROCCO: All right, we'll move on to some others, Commissioner. I would tell you that we have some real questions in terms of the validity, the reliability, the quality -- all of the issues that we brought forth, and whether the Department really should be using its resources elsewhere at this time. I think that's something that we'll be looking to hear more of from you and your staff, in terms of some of the issues that were brought up today, to kind of update us as we move along through this process. Because we'll be the people on the frontline, basically, that have to face our constituents each and every day who will be screaming bloody murder, unless everything goes pretty much in a good direction. That's difficult to determine. I, for one, think that there's a lot of quicksand in here that we will see in the year ahead.

I would like to thank you and Leo for being with us and for answering all of our questions. As I indicated originally, this is not partisan. It's not a witch hunt, or anything. We're just trying to find ways to make certain that the students in this State get the best possible testing program we can get for them.

COMMISSIONER FITZGERALD: I understand that. Thank you.

ASSEMBLYMAN ROCCO: Thank you.

ASSISTANT COMMISSIONER KLAGHOLZ: Thank you.

ASSEMBLYMAN ROCCO: Thank you.

We have a number of additional speakers, and we're going to go to that list. I would like to put into the record, David, the letter from Commissioner Ellis, and the letter from Tom Corcoran in regard to this issue.

Patricia Wang is next.

P A T R I C I A W A N G I V E R S O N: Assemblyman Rocco, may I wait until some other people speak?

ASSEMBLYMAN ROCCO: Sure.

MS. IVERSON: Thank you.

ASSEMBLYMAN ROCCO: Ellen Oppenheimer.

E L L E N O P P E N H E I M E R: That is I.

ASSEMBLYMAN ROCCO: Even though our Committee has dwindled somewhat, everything will be going into the record and will be transcribed. Ellen, I'm sorry.

MS. OPPENHEIMER: I'm here to talk to you about the impact of the early warning test in regard to my child. I have studied the test since 1990 with concern for the future of my son.

ASSEMBLYMAN ROCCO: Excuse me, just one second.

Commissioner, is there going to be anyone from your staff here? There are a number of parents here that are going to be talking about their individual situations.

COMMISSIONER FITZGERALD: Yes, Dr. Klagholz will stay.

ASSEMBLYMAN ROCCO: Okay. Ellen?

MS. OPPENHEIMER: I had some association with the test when I was teaching in the public school. Mostly the HSPT was being reviewed at that point in time. When my son was going to be tested on the early warning test, that school year, I had decided to take a real careful look at the test. I had requested from Dr. Masonis' office a copy of the previous test -- the 1991 test. Studying the test was of great interest to me. I had found that the test had appeared to be written with a process of two agendas. The math test and the reading portion of the test looked quite different.

We had decided that we would go over the 1991 test with our son. We had hired a tutor at the time, Mr. Erwin Oser, who had advertised in a local newspaper that he was an expert in the HSPT and early warning test, and he had turned out to be an employee of the State Department of Education. We worked on the test. My son took the test. We received his results on June 30. He did well on the math portion. He did well on the writing portion. He did not do well on the reading portion.

I immediately called my local school system, which is Manalapan. They had gone on their extended break for the July 4 weekend, so there was no one to speak to there. In turn, I had called the State Department and had spoken to Dr. Masonis. I had explained to him that I was most concerned about the score because I have now an 8th grader who is going into high school, and I wanted to make sure he was properly placed.

I told him at the placement procedures in February, he was placed in a college track academic program. I needed to know what the magnitude of the score meant in his programming. He told me that with the score at hand-- Not having my son's profile in front of him, that I should go back to the local and get ahold of them, and have them review the issue. I tried to do that. The local person, who was the reading specialist at that time, was Mrs. Schrader. She told me that she did not have the test booklet and that I would have to go back to Dr. Masonis. In turn, I had had a number of materials, including CAT scores, individual testing from the University of Pennsylvania, that had not been consistent with the performance of my son on the early warning test reading portion.

My basic concern was to where to properly place him? If he needed remediation, we wanted to provide it. If there was a question, I wanted to know which way to go.

ASSEMBLYMAN ROCCO: So Penn did some what, reading inventory?

MS. OPPENHEIMER: They did an entire battery for us, and that was approximately seven months prior to the EWT.

ASSEMBLYMAN ROCCO: Were those scores higher or lower?

MS. OPPENHEIMER: Much higher. Much higher. The individualized testing which is validated, which is reliable, which has been in the psychological venue for years, did indicate that he was above average, and certainly capable of moving forward in any direction he so chose.

ASSEMBLYMAN ROCCO: So you felt much more confident with the test that had been utilized and had been around for a while, as opposed to something totally new and--

MS. OPPENHEIMER: And also his academic performance as well as his CAT score.

ASSEMBLYMAN ROCCO: Historically, that had proven to be more correlated with the Penn test -- the U of P test.

MS. OPPENHEIMER: Yes. Actually the Penn test on the individual basis are far better tests. Any time you take a child and test individually, you get a far better evaluation. Those tests had indicated that he did far better than what he had achieved on the CAT test, and certainly far better than what he achieved on the early warning test.

I asked Dr. Masonis, about at that point, to help me. I said, "May I review this with you all? I have all of this data, and I'd like to sit down with someone to be able to make a pragmatic decision." He told me, again, that I would have to go to the local. The local told me, again, that they did not have the test booklet.

I had decided to place my child out of the public school system. He had asked to be placed into a private high school. We have done that. We researched the high school about the scores. We gave them all of the documentation, and we told the high school that based on the comments from Dr.

Masonis and Mrs. Schrader, my son would be a candidate for remediation, looking at the early warning test. The school had said to us that they would err to the conservative and place him in the remedial program, and watch him in order to decide whether or not we should continue that placement.

But because they did not know him-- And one of the faults of the early warning testing program is that when you are tested in the 8th grade and move on to a new system, that new system does not know the child, and all of the files that you send over, still are not the child. The reality is they did not know this child, and would err to the conservative, as opposed to saying, "Well, let's just see how it goes."

We had him placed in the remedial program. Within approximately 10 weeks we had a progress report from the staff at the high school. I spoke with the remedial teacher who said this child did not belong in her class. I said, "Well, what about the HSPT? Those are my concerns." She said, "Well, if you'd like to keep him here, we can, but I really feel that he would do fine elsewhere." I decided to keep him, just to see how things would continue. We kept him there and, again, we were concerned that this was not exactly the kind of program he needed.

We started to investigate our request again for Adam's test folder through Dr. Masonis' office in February, when all of the press releases came out with regard to the concerns of the early warning test. I called Dr. Masonis, and I said to him, "Now I am no longer requesting my son's test folder. I am telling you that I want that test folder. I want to review it. I need to make sure I am doing the right thing." He said he would have Dr. Roberts contact me, and, in fact, she did. She contacted me and said that she would have a date and time that we might meet to review the test folder.

Twenty-two days later, four phone calls to four different Assemblypersons, a phone call to the Governor, an

unreturned phone call to Dr. Roberts, constant pursual of this issue, we were able to get a time and date where Dr. Roberts and I could meet. Only there was a glitch; there was no longer a test folder. There was a data base that would indicate what my son's responses were, but I would actually not have his test folder. It seems that all of the test folders -- the actual documents that the children filled out -- had been destroyed.

ASSEMBLYMAN ROCCO: For that district or--

MS. OPPENHEIMER: For the entire group. All the tests are gone.

ASSEMBLYMAN ROCCO: For the State?

MS. OPPENHEIMER: You will have to ask Dr. Fitzgerald that. But the tests are gone. So now I was looking at a suspect test with dubious data based testing results. Needless to say, we were most concerned.

I met with Dr. Roberts. I went back to my school system. I had gone back to the private school and the private school had said to me, "We do not believe your child belongs in the remedial program based on his entire profile. It is apparent that you have been misguided, and in your zeal to do the right thing, we have not truly worked to the best interest of your son." They then recommended that we take him out of the remedial classes, and, in fact, that's what we had done.

The concerns that I have could have been completely mitigated by simply Dr. Masonis' efforts in June to get that test booklet and review the data. Instead, he had indicated to me that the early warning test did indicate that my son was a candidate for remediation. I had told him about the testing that I had. I asked to have all of this reviewed. No one picked up the gauntlet. Instead, my son was placed in a remedial program in lieu of being able to take a science program in Pascal.

I was particularly concerned about Dr. Fitzgerald's comments with regard to the fact that the children who were

placed in remedial programs, these programs really do no harm. My concern with that statement is, I have a question with regard to what price present and future gains, as well as a child's self-esteem. Those classes do harm. They're not meant to do harm, but if you are misplaced in those classes, you could be doing other productive kinds of purposeful, educational pursuits. So, I was in complete disagreement with that statement.

I'm also concerned with the fact that we were never, ever told that this test was in question. Dr. Masonis knew at the point I spoke to him in June, that there was concern about the test. He knew he had a concerned parent on the phone. He also knew that the parent was asking for assistance and review of the data, so that the parent could make a pragmatic decision for her child that was not afforded to this parent, even though there was concern with regard to the test itself.

ASSEMBLYMAN ROCCO: Let me ask somebody in the Department. The test documents were destroyed, is that correct?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes, and we recorded that in our investigation report on the final page.

ASSEMBLYMAN ROCCO: Page 15.

MR. HESPE (Committee Aide): Mr. Klagholz, could you come forward so we can pick you up, please?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes.

MR. HESPE: And then, can we repeat the question and the answer?

ASSEMBLYMAN ROCCO: The question I have to the Department in following Ellen's testimony, basically, is: Were the test documents that students responded to, destroyed?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes, Assemblyman, by the scoring subcontractor. It involved the multiple choice sections rather than the writing and open-ended sections of the data base that was referred to. Copies were being kept electronically, and can be reproduced. So the data are there,

but what the subcontractor did do, was destroy the actual folders inadvertently.

ASSEMBLYMAN ROCCO: Okay, so you're saying that you never authorized the subcontractor to destroy the--

ASSISTANT COMMISSIONER KLAGHOLZ: On the contrary. The contract requires them to maintain the folders for a year.

ASSEMBLYMAN ROCCO: And from the document here, it indicates the warehouse supervisor thought he received a verbal approval to destroy the documents, and destroyed them?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes. That section is a quote from a letter that the contractor wrote to me.

ASSEMBLYMAN ROCCO: Thank you.

ASSISTANT COMMISSIONER KLAGHOLZ: You're welcome.

MR. HESPE: Has the procedure for handling the documents changed since this mishap?

ASSISTANT COMMISSIONER KLAGHOLZ: No. It's the same contract that will require them to maintain it for a year.

MR. HESPE: No, I mean as to verbal-- He said he got a verbal approval and then he destroyed them. Has that been changed at all to prevent this?

ASSISTANT COMMISSIONER KLAGHOLZ: Yes, the contractor has sent a description of procedures they've instituted to avoid that.

ASSEMBLYMAN ROCCO: Does that include written--

ASSISTANT COMMISSIONER KLAGHOLZ: Yes.

ASSEMBLYMAN ROCCO: --written from the Department?

ASSISTANT COMMISSIONER KLAGHOLZ: From the contractor to us.

ASSEMBLYMAN ROCCO: No, I guess the question is-- Before the contractor would destroy a document, I'm suggesting that he should have written approval from the Department as opposed to verbal approval.

ASSISTANT COMMISSIONER KLAGHOLZ: Absolutely. That was in our contract. What there was, essentially, is not

inadequate procedures, but a violation of the contract and of the procedures that were established.

MR. HESPE: Is the same contractor being used after this violation of procedure?

ASSISTANT COMMISSIONER KLAGHOLZ: The contract has been reoffered, and it's currently in the final stages of the bid process.

ASSEMBLYMAN ROCCO: That was a pretty grievous error there.

ASSISTANT COMMISSIONER KLAGHOLZ: Indeed, it was.

ASSEMBLYMAN ROCCO: I would suggest the Department take steps to make certain that doesn't happen again.

ASSISTANT COMMISSIONER KLAGHOLZ: We will.

ASSEMBLYMAN ROCCO: Because, obviously in the case of Ms. Openheimer here, there's no way to get that student's document to show just how each and every question was responded to.

ASSISTANT COMMISSIONER KLAGHOLZ: Exactly.

MS. OPPENHEIMER: I'm particularly concerned about all of this. I also spoke to Dr. Susan Phillips, the person who wrote a report with regard to the early warning tests.

ASSEMBLYMAN ROCCO: She's from Michigan?

MS. OPPENHEIMER: Michigan State University. She had indicated to me that she could not discuss the early warning test, but I asked her if she would listen to my point of view and asked her if she thought it was viable to come and address this Assembly. She was quite positive that it was important for you to understand that this test, from my point of view, was not ready for these children; that this test has had a number of excuses, that people are constantly excusing this test, but the bottom line is that these children were harmed. It wasn't just 5000 children; there were 34,000 children involved in not passing this test. This is a travesty, and this Committee has made some very clear comment on that concern

in terms of the future. But we now have a group of children whose parents have had to sign waivers in order to keep them out of remediation, based on the early warning test. We have a group of children who are in remediation. We have a group of children who have been pulled out of remediation. Those remediation teachers are paid in full for the year, even if that child is no longer participating in that program. That's another cost factor, as well.

ASSEMBLYMAN ROCCO: You heard the testimony this morning. What's your reaction to some of the questions and responses by the Department?

MS. OPPENHEIMER: Well, I have to tell you candidly that I am very concerned about the personnel in the Department. I'm concerned only on my own personal level. We were not given disclosure. We were not afforded our due process in order to make a pragmatic decision on the very child that these people are supposed to be responsible for.

ASSEMBLYMAN ROCCO: In your comments with Dr. Phillips from Michigan State, do you think that the Department responded they have little or no concern about the test for the future? Did that seem to be in concert with what Dr. Phillips and you had discussed?

MS. OPPENHEIMER: Dr. Phillips was limited in her comment. My concerns were many of the ones that you had addressed. We have a State teachers' test that we give to our teachers in order for them to get a minimum score so that they can teach in this State. We have testing for the children so that we know where we stand. Now we have the early warning test, the HSPT, and you are concerned; and rightfully so. Your concerns in total are, from my point of view, thoughtful, and provocative as well. I think you've made your point. But the reality is that there are a number of children that have been hurt presently, and are hurting right now.

ASSEMBLYMAN ROCCO: Do you think something more should be done for the students who were misplaced, or who did not pass this exam?

MS. OPPENHEIMER: Yes, I do. Clearly, notification is not being given properly to the families. Certainly they are going to the superintendents, to the principals, but that is not coming down to the families.

Secondly, I think in terms of the children, I think those '91 scores and the '92 scores do not appear to be solid. I think they should be removed from the children's profile, their academic profile. They can only hurt; they can't help.

ASSEMBLYMAN ROCCO: Can I ask the Department, again, are these being recorded in a cum(phonetic) folder or in any other document that follows that child through the system?

ASSISTANT COMMISSIONER KLAGHOLZ: I think that they are, Assemblyman, but I'm not absolutely positive of that, and I'm sorry.

ASSEMBLYMAN ROCCO: Could you make certain, as a request, that they are not, since there is a real question about that validity?

ASSISTANT COMMISSIONER KLAGHOLZ: Okay.

ASSEMBLYMAN ROCCO: And I think a parent has a perfect right to have a concern here, in that regard.

MS. OPPENHEIMER: And, again, my final concern would be what your concern was, and I thought it was quite insightful. This effort may be misplaced. The Department of Education needs to focus on child centered core curriculum that can be carefully disseminated throughout the State, so that the State can educate our children properly. We don't need another standardized test to tell us we're not doing a good job.

ASSEMBLYMAN ROCCO: I don't know how much more I can say on the issue. I kind of hit it in four or five different directions. The Department is an arm of the administration, and so they must move ahead. We can pass legislation to

override what the Department does, and then we, of course, would have to have the override after that. So, I think this Committee was very strong in their feelings in regard to it, and my guess is the Department will ignore it and move ahead as they have, and will continue utilizing the in-house exam. That's my guess, which will probably be 95 percent correct.

MS. OPPENHEIMER: I just want to remind those who are working through this test and who have to represent their constituency as well, that the very same children who do poorly on the HSPT, in a year-and-a-half will be their voting constituency. That needs to be reflected upon with regard to that issue, because I can tell you from a 14 year-old's point of view, that the statements made by Dr. Fitzgerald with regard to the remediation as not being a bad thing, were catalytic in my home.

ASSEMBLYMAN ROCCO: Certainly.

MS. OPPENHEIMER: And that's the message that he wishes to--

ASSEMBLYMAN ROCCO: And it would have been catalytic in my home. There's no doubt about it.

MS. OPPENHEIMER: And that is from a child who did not do well on the reading portion of the EWT. So I just want to make clear to you that there needs to be representation for the child, as well as for the parent. The bottom line: The person who is really being left to carry this burden is our child.

ASSEMBLYMAN ROCCO: That's true, and you as a parent have the ultimate responsibility in that regard. We appreciate your comments because I think they focus on the issue much better than we can, simply looking at it from our perspective. You bring a perspective as a parent which is very critical to this whole issue.

I would basically say to you that my fear is the Department has a pride of authorship. They want to be the first-- Forget about whether it's right or wrong. They just

want to be the first, or the leader of the pack, even if the pack is going in the wrong direction. I think they're trying to prove a point here that the millions of dollars that have already been spent haven't been wasted, when, in fact, we all know they have been wasted and much more will be wasted.

So I don't have a very good feeling about where the Department is going from this point, in this regard. I don't think this meeting this morning will do much more than to have them kind of put up defenses in various positions. But in truly looking at some of the issues that were brought to the fore, I don't see it happening.

MS. OPPENHEIMER: It is most unfortunate that these are the very same people who could have made a difference in terms of my son and so many others. That is an unfortunate reality. But again, you said the ultimate responsibility belongs to the parents.

ASSEMBLYMAN ROCCO: It certainly does.

MS. OPPENHEIMER: These parents. My two sons have been taken out -- have been placed in the private venue, as opposed to the public venue. I think that this situation clearly exhibits why.

Thank you for your time.

ASSEMBLYMAN ROCCO: Thank you.

Joe Rosenstein, New Jersey Math Coalition.

J O S E P H G. R O S E N S T E I N, Ph.D.: Good afternoon, or good morning.

ASSEMBLYMAN ROCCO: Good afternoon, Joe. By the way, before Joe starts, we are going to go through the rest of the meeting, for the people who are going to testify here. So if anyone has to have lunch, feel free to leave us and come back, if you so desire.

DR. ROSENSTEIN: My Chairman and Committee members, let me introduce myself. My name is Joseph G. Rosenstein, and I am in the Mathematics Department of Rutgers University, New

Brunswick. I am here today as Director of the New Jersey Mathematics Coalition, which is a partnership of the education, business, public policy, and public sectors of New Jersey, all working together to improve the mathematics education of our children.

I am here to tell you that those who are actively involved in improving mathematics education in the State strongly support the directions taken in the mathematics portion of the High School Proficiency Test and the early warning test.

I am not used to this kind of a forum, so I am not sure that what I say will come out the way I would like it to. I also am very aware that many things that people said today I would like to respond to, but my time is obviously going to be limited, so I will be relatively brief. I would like to take a few--

ASSEMBLYMAN ROCCO: Joseph, that is the wonderful part of America. You can have your position, state your position, and certainly we will take it into consideration.

DR. ROSENSTEIN: I would like, however, a few moments to explain why I make that statement. The basic fact, whether we like it or not, is that teachers teach to whatever the test is, and that districts write curriculum based on what they think the test is. That makes it particularly important that we test what we really value; that the tasks that we give children on our assessments reflect the tasks that we want them to perform.

For a number of years, standardized tests, particularly in this country, have focused on asking multiple choice questions which measure rote learning, and schools have responded by focusing on rote learning. Earlier, Assemblyman Doria mentioned that the only guideline in the legislation was that students should -- that as a result of this examination, or the tests that are used, it should be measured whether

students can function politically, economically, and socially in general society. I submit to you that none of the standardized tests you have talked about have that as their goal at all. That is not their goal. Their goal is to measure what can be called "school learning," and they measure it in a rote way. As a result, many of the students are turned off, to mathematics, for example. There is almost universal agreement that rote learning, the kind that is measured by the tests you are talking about, is not enough for citizens in the next century, and it will not enable us to achieve what former President Bush set as one of his goals, to be first in the world in mathematics and science by the year 2000.

Now, that goal may be too ambitious, but there is no doubt that we can do much better, and there is a national consensus on how to improve where we are going. The pack is not going in the wrong direction, Mr. Rocco. The pack is going in a very positive direction, and New Jersey being a leader of that pack is a positive thing. I have no stake in whether the Department of Education is right or not in this but it is moving, in this case, I think in a very healthy direction, and I think that must be reassured.

ASSEMBLYMAN ROCCO: Joe, don't you think we have to look at the direction of your Coalition versus a testing instrument, basically, to test, as opposed to-- You know, I guess I am concerned about confusing what the Coalition is doing versus what is occurring in the testing itself. So let's not--

DR. ROSENSTEIN: What the Coalition is doing is completely separate from the test. We are not an advocate for the test, and we have not made the test. But I would like to give some examples as to why this assessment is moving in the right direction.

The standards which these assessments are addressing, even if imperfect at the moment, are standards which have been

approved by all the national organizations: the National Organization of Teachers and Supervisors of Mathematics. They have the stamp of approval of Lamar Alexander, who is the previous Secretary of Education in the Bush administration. But I would like to give some examples as to why these tests are doing something which is very important.

Let me take one example: One of the open-ended problems on the HSPT -- I read the test, so that is where I get it from -- involves a fast food manager who has to use information about hourly gross receipts to schedule her employees so that their time is not wasted. Now, is that a skill that we want our students to have? I submit that it is. I submit that it is a skill that we want all high school graduates to have, not just college-bound ones. I think that is something where if we do not require students to be able to answer that kind of question, we are sacrificing them.

ASSEMBLYMAN ROCCO: But, Joe, how is that different from a word problem? Do you know what I mean? Where do you get the difference there?

DR. ROSENSTEIN: Let me try to say a little bit about the difference: In a word problem, all of the data is given in a very compact sentence. I say to you, "You have this amount of money, and you want to divide it among this number of people. How much does each one have?" Okay? That is a word problem. Word problems-- In your lifetime, in your career, whatever your career was besides the Assembly, or even in this career in the Assembly, no one ever asks you a question in a one-sentence format. You have difficult problems to wrestle with.

The manager of a fast food restaurant has a difficult problem to wrestle with. He has to figure out how many employees to have at each time of the day, and he has to figure out how to do that. That is not a word problem; that is what one of our New Jersey scientists called a "world problem."

ASSEMBLYMAN ROCCO: Why wouldn't it be a word problem? You can have a word problem that goes past one paragraph. A word problem is not restricted to one paragraph.

DR. ROSENSTEIN: But generally they do. Those on the standardized tests are one paragraph word problems. They are supposed to be simple to understand and simple to do.

ASSEMBLYMAN ROCCO: No, not always.

DR. ROSENSTEIN: But they do not reflect the kind of tasks which students have to do as they are students, whether they need that for their careers or for their education, whether they need that as citizens or as consumers. They need to be able to look at information and be able to process it for themselves and be able to figure out what to do with it. That is the kind of skill, higher order problem solving skills, that are recommended by all the national reports -- nonpartisan all national reports. Those kinds of skills are attempted to be reflected in this test.

I think the Commissioner said that perhaps five years from now there may be many such assessments, but now there really aren't any. If we wait, we could obviously do that, but if we wait, then we shortchange our students for another few years until we are ready to have such an assessment.

Let me give another very simple example.

ASSEMBLYMAN ROCCO: Be careful on examples. I have seen eight million of these examples. Okay? Be ready to defend here.

DR. ROSENSTEIN: Okay, all right. On standardized tests, you typically get very simple addition problems. On the HSPT, for example, students are provided with a calorie table and asked which of several meals will come in at fewer than 800 calories, for example. A slightly more complicated problem. It requires some thought; it requires processing information; it requires what are called "higher order of thinking skills." If we want those kinds of thinking skills to be in our curriculum, they have to be included in assessment.

If, as has been the case for many years of our standardized tests, the only kinds of questions we ask are rote questions -- what is 6 percent of eight? -- that is all the students will learn, because that is all that they will be taught. If we want to go beyond that, our assessments have to go beyond that. What schools set their curriculums up to do is to achieve the assessment. If we want our students to achieve higher order skills, then we have to set that in the assessment.

Unfortunately, there is no assessment now which does that. Okay? The only ones that are presently there are the ones which are being developed by the State. I don't know whether New Jersey is way ahead of all the others, but certainly what we have reflects those kinds of goals, those kinds of standards for our students. If we don't have a test which says, "These are high standards, and we want our students to meet them," then the schools will not meet them.

Let me say a word about remediation. Obviously there are going to be problems for several years. I can't address those because I don't know the magnitude of those, but there will obviously be problems. We have heard a good deal about them. Our goal is, not that there should be more and more remediation, but that there should be less and less. If, in fact, the message comes across clearly to schools that this is the kind of assessment which we expect the students to succeed at, and if the message is given to them that all students -- perhaps a few exceptions, but all students can achieve that level of understanding of mathematics, if those are the case, then the schools will gear themselves up to address that. They will have to. No district will be willing to be the object of public ridicule because it has not. Each school will, in fact, do what is necessary, and with the proper support the staff will do what is necessary to have those students achieve the necessary skills.

Remediation is something that we do not need. We have to get away from the idea that we should have perpetual remediation. The only way to do that is to ensure that from the outset very clear standards are set, and from the outset it is understood that those are the standards that we measure. We must hold all students to those standards. It is not enough to say that some students can do it and some students can't; too bad about those, forget it. We really must insist for our society that we address the needs of all students. And if some students do not have the resources to meet those needs, then we must provide adequate resources for that to happen.

The New Jersey Math Coalition is committed to the idea of math for all. That is to say that all students can achieve in mathematics. There are many students out there who believe that they cannot do mathematics. There was one young lady in one of the 30 urban districts who told me, "Math doesn't run in my family," as if it were somehow a genetic -- the ability to do basic mathematics is genetically determined. Only in this country do people believe that. We have a big job ahead of us to convey to people that they can achieve in mathematics; that their children, that our children can achieve in mathematics.

Having standards and having an assessment which reflects those kinds of standards can go a long way. In the end, it will be a less expensive way of doing it than by piling remediation upon remediation and having generation after generation graduate from school without having the skills they will need for the next century. Those skills are much more than we need now. As some people pointed out looking at this test, "That is not what we were asked in school." Well, that is not what these kids are going to need when they enter the job market a few years from now, and we will be shortchanging them if we continue as if the skills that are needed in the next century are the same as the skills which were needed in the last one.

Maybe I should stop there.

ASSEMBLYMAN ROCCO: Okay, Joe. Thank you.

Patricia Clark Kenschaft.

P A T R I C I A C L A R K K E N S C H A F T, Ph.D.: I have a written talk, but after listening for three hours I really have to respond.

ASSEMBLYMAN ROCCO: Please.

DR. KENSCHAFT: I am a Professor of Mathematics with a doctorate in Pure Mathematics from the University of Pennsylvania. There are roughly 30,000 if us, including Joe and me, in the entire country with doctorates in mathematics. That compares to 50,000 lawyers in New Jersey. We sometimes feel misrepresented and misunderstood, and sometimes we don't act the way we should when we feel that way.

But the honest truth is that mathematics education in this country, what is taught in the schools in the name of mathematics, has gotten far, far out of line with the international understanding of mathematics over the last centruy. It has been a century of divergence where we have been talking less and less with the schools, and they less and less with us. It is very sad for all concerned.

The current national tests are measuring what the schools call mathematics, which is really very, very different from what I understand to be mathematics. I think mathematics, as the international community understands it, is an important thing to be taught in United States schools. I assume you know that appalling test where the top half of the Japanese 18-year-olds were the same as the top 5 percent of our 18-year-olds. It is not genes; it is the way the schools are run.

Now, the last dozen semiannual--

ASSEMBLYMAN ROCCO: The real question is: What instrument was used?

DR. KENSCHAFT: Pardon?

ASSEMBLYMAN ROCCO: What instrument was used?

DR. KENSCHAFT: We had some international consensus on those instruments. For example, in other countries, algebra is taught between the 5th and the 7th grades, to almost everybody -- 90 percent to 95 percent of the kids. In our country, it is taught in 9th grade to 40 percent of the kids, and that is four years retarded. We have the same genes as they do in Nigeria and Japan and Europe. It is not genes.

ASSEMBLYMAN ROCCO: When they measure the international community, what tests do they use?

DR. KENSCHAFT: Do you want me to read it to you now? Professor Stephenson -- where is he from? Maryland, I think. I have it at home, if you really want the test, and the write-ups of the test. Professor Stephenson has devised tests, in cooperation with people from other countries. But we are devising mathematics tests which are much more in line with the ones that the Department is writing, than the standard--

ASSEMBLYMAN ROCCO: I guess my question is: To back it into New Jersey as if we say we are not competing internationally, and as the data that you gave earlier about the Japanese children being able to score much higher than ours-- Based on what criteria?

DR. KENSCHAFT: There are a variety of different tests. They have been done in three different cities, in -- was it-- Gee, I haven't looked at this recently. We're talking months. But dozens of different schools--

ASSEMBLYMAN ROCCO: I am not saying they are inaccurate. I'm saying, you know, they have been validated over a good period of time, I would assume.

DR. KENSCHAFT: Right. They have been done--

ASSEMBLYMAN ROCCO: That's the problem with the New Jersey tests. Not enough validation. That is the problem we face.

DR. KENSCHAFT: But they are more in line-- See, the national test that you keep referring to as being validated-- They are testing something that is of no interest to me. I have been going to--

ASSEMBLYMAN ROCCO: It doesn't matter whether it is of interest to you. What matters is whether or not it is of interest to the high schools, the colleges, those who must make judgments based on those tests.

DR. KENSCHAFT: Well, that is an interesting statement. I want you to think about it. Is it really getting into college and so on that is the goal of the K through 12 education? Or is being a useful citizen and being able to know some internationally respected subject matter what we want to--

For example, in every other country you have to pass--

ASSEMBLYMAN ROCCO: Well, you know, we can philosophically debate from here until next July.

DR. KENSCHAFT: Right. Okay, let's move on to some--

ASSEMBLYMAN ROCCO: The question is: What are the instruments?

DR. KENSCHAFT: I want to say--

ASSEMBLYMAN ROCCO: How effective are they? Is the New Jersey variety truly valid and reliable? Those are the questions.

DR. KENSCHAFT: There are two other questions.

ASSEMBLYMAN ROCCO: No. They are the only questions. Validity and reliability of the existing exam becomes the issue in the State of New Jersey. There is already great concern about that issue. That is why 5000 students -- and many more may well have been misplaced-- That is why you have to wonder how test exams get destroyed. That is why there is a great deal of concern by the individual from Michigan State -- Phillips -- with regard to the test -- validity and reliability.

DR. KENSCHAFT: But also, the subject matter. In every other country, you must pass calculus to get into

college, in every field. Now, we have a lot of educated people in this room, but I'll bet most of them didn't take calculus in high school. That is expected in every other country. How are we going to get incrementally from here to there?

My project-- Now I am slightly off the subject, but I think you will like this, so you'll want to listen to me. I have been working for the last five years with K through 4 elementary schoolteachers. I really think we need to help them. I have found them intelligent and very eager for help. However, they don't know fractions. They tend not to know areas. Many of them do not understand what subtraction is. These are important subject for these tests no matter what tests you have. If we don't help K through 4 teachers to learn mathematics, we're in trouble.

One of my protegees in the worst part of Newark -- this is the worst socioeconomic system -- her 3rd grade kids did score in the 70th percentile last year in the old-fashioned test. Yes. Newark rarely breaks the 40th percentile, and my protegee's kids got the 70th. The old-fashioned test, as one of you was saying -- and I don't think it was you, Assemblyman Rocco -- if we really teach mathematics, they will reflect. But as Joe was saying, and the woman before me, the teachers teach to the test. So, there you are. The teachers do teach to the test, so what tests we have really, really matter. Unless we strengthen the teachers to have the courage to teach real mathematics, in which case their children also will do well at the old-fashioned American test-- I would grant you that, but somehow we have to liberate the teachers.

Every speaker I have heard in front of teachers says, "When the tests are coming, all education stops for two weeks," and the teachers all nod, because the current tests are so far out of line with what we want education to be, that scoring high becomes the goal, instead of educating the children. It is so sad to watch children being needlessly destroyed.

I have worked with children enough in Newark to know that up until the age of 10 they can learn mathematics and learn it quickly from me and the teachers who have been working with my team. I will leave you some of my information. But we are about to be closed down, probably. We may be; at least I will be because the college needs me back again. If we don't strengthen the teachers, and if we don't get the tests back in line with what we understand to be mathematics, our children are doomed to economic disaster, which long-term is going to be just terrible for them, not to say us.

ASSEMBLYMAN ROCCO: Thank you, Patricia.

Dr. Ira Sweet.

I R A S W E E T, Ph.D.: My statement will only be about five minutes, and I hope you will indulge me.

My name is Dr. Ira Sweet, and I have been employed as a teacher, guidance counselor, and school psychologist in the New Jersey public schools for over 27 years. I have also been associated with Trenton State College, Kean College, and Ocean County College as an adjunct faculty member in psychology for over 20 years.

I appreciate the opportunity given to me to testify at this hearing, and it is a sad affair that I have to be here. When I marched with Martin Luther King in Montgomery, Alabama over 25 years ago, I was there to demonstrate for social justice and against institutionalized racism and discrimination. It is now 25 years later, and I must now speak out against elements of institutionalized discrimination which exist within the State Department of Education as it relates to testing procedures and practices.

Understanding the nature of this hearing, I should let you know beforehand that my remarks are related to testing problems in the public schools of New Jersey, but are not directly related to the current test being discussed. You are concerned about 5000 students being placed in remedial programs

because of inappropriate testing. I am concerned about students being classified as mentally retarded and placed in inappropriate special classes and programs because of inappropriate tests being used and supported by the State Department of Education. I will let you be the judge as to which problem is the most damaging to children.

ASSEMBLYMAN ROCCO: Dr. Sweet, you know, I want to give you as much latitude as possible. This hearing today is based on the testing, the 8th grade early warning test. Really, your other information will be placed in the record. We are having everything transcribed. We would prefer that you stay with the issue at hand.

DR. SWEET: I am asking you to expand the scope of your investigation to find out why personnel from the State Department of Education are knowingly allowing school districts to use an intelligence test that is not valid; that discriminates against handicapped children; and that is culturally biased and racially discriminatory.

ASSEMBLYMAN ROCCO: Which test is that, Dr. Sweet?

DR. SWEET: Let me finish. You have people testifying here about discriminatory testing in the State of New Jersey. I am talking to you people about children being allowed to be tested by a test -- and I will mention the test and show you the test -- that is classifying them as mentally retarded.

ASSEMBLYMAN ROCCO: What test are you talking about?

DR. SWEET: I am talking about the Slosson IQ Test.

ASSEMBLYMAN ROCCO: Slosson has been around for 30, 40, 50 years.

DR. SWEET: It's been around for a number of years, yes. Will you give me four more minutes?

ASSEMBLYMAN ROCCO: We don't want to deal with Slosson. Slosson's correlation with Wexler, I think, is pretty well defined.

DR. SWEET: How is it defined?

ASSEMBLYMAN ROCCO: Let's deal with the issue. I don't want to get into that debate. I want to deal with the issue today, which deals with the test -- the early warning test -- and the High School Proficiency Exam.

DR. SWEET: I will leave you my testimony. I can't testify today.

ASSEMBLYMAN ROCCO: Well, thank you.

DR. SWEET: I'm shocked that you will not let me testify. You had a person testifying here, a black woman, about discrimination.

ASSEMBLYMAN ROCCO: The meeting is adjourned.

DR. SWEET: All right.

ASSEMBLYMAN ROCCO: Oh, I'm sorry. Patricia Wang-Iverson.

MS. IVERSON: I first heard about this hearing through my association with the New Jersey Mathematics Coalition -- the March hearing, which was postponed. The April 19 hearing I heard about in my capacity as a member of the Board of the New Jersey State PTA. But right now I would like to speak as a private citizen; as the parent of a 12-year-old and a 10-year-old; and also as an immigrant who crossed three continents to come to the United States for the wonderful educational opportunities. So my interest in coming here is to support a collaborative effort on the part of all sectors to work together to do what is best for the students.

What I would like to do, rather than passing it to you after I speak, is to share some national material that I have gathered, and I would like to speak generally on testing. I have some extra copies if there are people in the audience who would also like to look at these. (holds up materials)

I would like to just speak generally. If you look at the "Statement of Principles," what is unique about this is, it is the first time the U.S. Department of Education and the National Science Foundation have collaborated to put out what I

feel is a very succinct publication. The full title is, "Statement of Principles on School Reform in Mathematics and Science." In fact, there is a section on student assessment:

"For purposes of accountability, states should develop new student assessments based on national content standards and state curriculum frameworks. These new assessments should test students' knowledge and understanding of mathematics and science in ways that are more complex and demanding than current tests."

The Assembly and the Senate should be thanked, because through your mandate the Department of Education, in fact, is creating standards in every discipline, isn't it? They are working on this. New Jersey is one of six states to be awarded a grant by the U.S. Department of Education to develop curriculum frameworks in mathematics. Again, all sectors are working together on this.

The other point, and this other publication is called-- It is a blueprint from the Mathematical Sciences Education Board which was created by the National Research Council. They have come out with a publication called, "Measuring Up," which is, again, looking at assessing children. I just want to point out one phrase: "Our children will be better off mathematically if state leaders demand measurement of what's worth learning, rather than just what's easy to measure."

What I would like to see for all children in New Jersey is what we are trying to provide for our children. But we know that we have an increasing number of children who don't have parental support, and their only hope is the school system. How can we help the school system provide what is best for these children?

In final summary, I come to testify on behalf of the EWT and the HSPT, and with the hope that we are all going to work together to try to improve them.

100

BEST COPY AVAILABLE

ASSEMBLYMAN ROCCO: Thank you very much, Patricia. I don't think anyone disagrees with your objectives at all.

MS. IVERSON: I'm sorry.

ASSEMBLYMAN ROCCO: I don't believe anyone disagrees with your objectives. We would all like to see that occur.

MS. IVERSON: Thank you for this opportunity, and thank you for your efforts.

ASSEMBLYMAN ROCCO: Thank you.

Bob, do you have anything before we close?

ASSEMBLYMAN MARTIN: No, thank you; nothing.

ASSEMBLYMAN ROCCO: This meeting is now adjourned, at 1:00.

(MEETING CONCLUDED)

APPENDIX

Eighth-Grade Early Warning Test

Findings and Recommendations

Mary Lee Fitzgerald
Commissioner

Leo Klagholz
Assistant Commissioner
Division of Academic Programs and Standards

New Jersey Department of Education
225 West State Street
CN 500
Trenton, NJ 08625

March 1993

STATE BOARD OF EDUCATION

ANNE S. DILLMAN..... Perth Amboy
President

ROBERT A. WOODRUFF..... Elmer
Vice President

S. DAVID BRANDT..... Cherry Hill

RONALD K. BUTCHER..... Pitman

MAUD DAHME..... Annandale

ORLANDO EDREIRA..... Elizabeth

BONNIE J. GREEN..... Bargaintown

JOHN T. KLAGHOLZ..... Seaside Park

CORINNE MULLEN..... Hoboken

NANCY SCHAENEN..... Madison

DEBORAH P. WOLFE..... Cranford
Representing the State Board of Higher Education

Mary Lee Fitzgerald, Commissioner of Education
Secretary

Edward D. Goldberg, Chancellor of Higher Education
Ex-Officio

2X

On February 22, 1993 the Department of Education shared information produced by the current study with three consultants. Each of these consultants is nationally recognized in the field of testing and testing policy: Dr. Stanley Bernknopf, director of testing for the Georgia Department of Education; Dr. Sylvia Johnson, professor of statistics and measurement at Howard University; and Dr. Stanley Rabinowitz, former director of the New Jersey Department of Education's testing program who now works at Far West Laboratories in San Francisco.

On February 26, 1993, senior management of the department conducted an all-day meeting with these individuals to review the available factual record and discuss its implications. Based on these discussions, the Department of Education offers the following findings.

<u>The development and implementation of the 8th-Grade EWT has been hindered during the past three years by tight timetables, insufficient budgetary support, questionable administrative decisions, instability in personnel and a lack of strong policy leadership in the Department of Education.</u>

The 8th-Grade EWT was scheduled by law to be used formally for its intended purpose in the school year that began in September 1990. Funds for development of the test were not appropriated until July 1, 1990, two months before the start of the school year and eight months before the test was to be formally administered to students.

During that period of initial development, which was also a period of transition in state government, literally all Department of Education managers associated with the test program left their positions -- including the Commissioner of Education, the Assistant Commissioner of Educational Programs, the Assistant Commissioner of General Academic Education, and the Director of the Bureau of Statewide Assessment.

The State Legislature reduced the Governor's FY 1991 budget request for the development and initial administration of the new tests. Testing program budget requests have also been reduced in subsequent years. Budget cuts are a primary reason that field tests and validation studies were not conducted for the 8th-Grade EWT prior to its implementation, and have not yet been completed for the HSPT-11. Budget reductions are also indirectly responsible for the loss of four staff positions in the department's Bureau of Statewide Assessment.

The idea of including open-ended questions on the HSPT-11, and therefore on the 8th-Grade EWT, is educationally worthy and consistent with the purposes of the new tests.

However, the decision made in Spring 1990 to accelerate the inclusion of open-ended questions in the tests was practically unsound. This decision increased the cost of developing both tests at a time when resources were

BEST COPY AVAILABLE

already insufficient and timelines were tight. The decision contributed to the elimination of field tests and validation studies from the plan for development of the 8th-Grade Test. The Department of Education's immediate inclusion of open-ended questions in all aspects of the test scoring and equating process may have outpaced the ability of scoring technology to handle such items.

The failure of upper management in the Department of Education to respond to or resolve issues, to set priorities and to provide policy direction contributed to the exercise of poor judgement by technical staff and supervisors. At best, the assessment bureau was required, through the inattentiveness of upper management, to make all technical decisions and all policy decisions. At the worst, staff might have been pressed in ways that encouraged the exercise of poor judgment. Lacking adequate direction and support, the technical staff failed to resolve the issue of the 1992 reading scores in an effective and timely way.

The New Jersey Eighth-Grade Early Warning Test is sound and valid.

The inadequacies described above produced several undesirable consequences. However, they did not nullify the soundness or validity of the 8th-grade test itself.

The 1991 test items were developed and selected using accepted procedures. The test was administered throughout the state in a timely and efficient manner. The effectiveness of each test question was analyzed, after the initial administration and before students' tests were scored, to remove any ineffective questions from the scoring process. This analysis involved several levels of review by statisticians and content specialists. The procedure used is one that duplicates, after the fact, essential elements of the field tests and validation studies that ordinarily are conducted in advance.

As a result of this analysis, only two questions had to be removed from the scoring process in 1991. Both of these items were eliminated from the reading test. All of the experts with whom the department consulted agreed that the results of item analyses indicate that the 1991 Eighth-Grade Early Warning Test and, in particular, the reading portion of the test was well within acceptable limits of soundness and validity.

The problem that occurred in 1992 did not involve any portions of the HSPT-11 or the writing and mathematics sections of the EWT. It involved one section of one test.

Within the requirements of law, the Department of Education should have explored alternative approaches to standard-setting in 1991 and 1992. In particular, the department should not have attempted to equate 1992 test results with those attained in 1991.

4X

Problems that occurred in 1992 with respect to the reading section of the EWT were problems of scaling and equating, not problems of test validity.

Equating is a powerful statistical tool. However, it is most effective when applied to tests that are comparable and stable. The 1992 and 1991 versions of the Eighth-Grade Test were not comparable or stable.

As noted above, the test was designed initially using accepted methods and its questions were screened through item analysis. Yet, this positive result was achieved under stringent time and resource contraints and without benefit of advance field testing and validation studies. Therefore, there was room for improving the test in 1992.

Ideally, the Eighth-Grade Test should have been developed and implemented in the same way as the HSPT-11. The test items should have been field tested and validated in advance of formal implementation under the law. The test should have been pilot-tested and refined over a one-year or two-year period. District educators might have been provided an opportunity to become used to the test and its results, and to plan ways of making programmatic adjustments to enable students to meet its standards. During the pilot period, they might have been provided extensive advance training in the interpretation and use of test results. Test standards might have been represented in the first one or two years, not only as <u>single</u> indicators of student needs, but also as very broad indicators. The setting of a cutscore and equating of year-to-year results might have been delayed until a degree of stability had been achieved.

Since deadlines did not allow these preliminary measures, the actual implementation of the test in 1991 and 1992 might have been treated, to the extent possible under the law, as a pilot effort. In some respects, this was done even as the test was being formally implemented.

The use of performance bands was emphasized as a means of interpreting test results in the first year. Given the advantage of a second year, the Department of Education and test advisory committees attempted to refine and improve the reading section of the EWT for 1992. This effort was commendable and correct from both an educational perspective and a measurement perspective. However, it also changed the test and rendered the 1991 and 1992 versions less comparable than they otherwise would have been.

The report that Dr. Susan Phillips submitted to the department in 1992 describes some of the ways in which the two test versions were not fully comparable. In addition, different procedures were used to establish cutscores for 1991 and 1992. Further, as Dr. Phillips's indicated and the department's three consultants confirmed, the effects of open-ended questions on equating are not fully understood.

In this context, the department expected too much in relying on equating to dissolve all the differences between the two test versions that resulted from a healthy early evolution of the reading test. The attempt to equate 1992 and 1991 results using highly precise statistical methodology was an overstatement of the stability the fundamentally sound, but evolving Eight-Grade Test.

The Department of Education cannot determine with assurity which 1992 equated cutscore, of the several produced, is the "correct" one.

Various 1992 cutscores resulted from at least four different equating runs:

1. sample-based without anchor form four
2. sample-based with anchor form four
3. population-based without anchor form four
4. population-based with anchor form four

The different cutscores resulting from the first three procedures are within a one-point range. Only the fourth procedure produced a cutscore two points lower than the one initially assigned.

Superficially, the fourth procedure (population-based with form four) seems best because it includes all students and all portions of the anchor test. However, population-based equating is very unsual, and the department's consultants confirm CTB's contention that there is sufficient amount of inherent error to account for a two-point range in cutscores when equating is done repeatedly under different assumptions.

The fourth procedure also has superficial appeal because it produced a more "believable" result -- no decline in student performance from 1991 to 1992. However, if the tests are not comparable, then neither are the results. The results simply should not have been equated in the formative years of the test.

During Summer and Fall 1992, the Department of Education was ineffective in resolving the problem of an apparent nine percent decline on the reading test. Nevertheless, if districts followed the department's guidelines for use of test results, then it is likely that few students were "unfairly" provided remedial assistance.

Districts, schools and teachers must be responsible for educating students to meet the HSPT standard, for continually assessing individual students' progress toward meeting that standard, and for providing students whatever special assistance they need. The 8th-Grade EWT is a reasonable means by which the state can help districts to identify students who may need special assistance. Clearly, the state can systematically align the content of the 8th-Grade EWT with that of HSPT-11.

However, a single cutscore on an 8th-grade test may not in every case predict with precision which students will eventually pass or fail a test of 11th-grade skills. Nor is it necessarily true that the best course of action in every instance is to place marginal students in a separate remedial course. Some marginal students might benefit more from special and/or improved instruction in developmental courses. In the longer run, districts might also aid marginal students by considering more fundamental improvements in the broader instructional program.

It would be unfortunate if the results of the 8th-Grade Test were to become a means of discouraging districts from assuming responsibility for making educational decisions and being accountable for those decisions.

For that reason, the department has consistently urged districts to:

use results of the 8th-Grade EWT as the primary indicator of student needs, not the only indicator;

consider other indicators in identify students who need assistance and those who do not;

avoid assuming automatically that every student who falls below the state cutscore needs special help, or that every student who surpasses the cutscore has no need for assistance;

give thoughtful consideration to the question of how best to meet the needs of marginal students, and not to assume that a placement in remedial course is always best; and

avoid removing ninth-grade students from developmental courses in order to place them in remediation (NOTE: The testing law, N.J.S.A. 18A:7C-6.2 specifically identifies after-school, weekend and summer programs as being among the appropriate vehicles for providing extra assistance).

All of the students identified as being within the two-point band of 1992 EWT cutscores clearly are marginal. Under state guidelines, districts should not have relied soley on the state cutscore in identifying students for remediation, and they should have considered alternative means of providing assistance. No student should have been removed from developmental courses.

Further, the Bureau of Statewide Assessment, in cooperation with CTB Macmillian-McGraw Hill, conducted special workshops in September 1992 to inform districts of the scoring problem on the reading portion of the test. Districts were urged to be particularly thoughtful in their use of reading test results.

CTB Macmillian-McGraw Hill made two errors in 1992, neither of which was an equating error. In Summer and Fall, 1992, CTB also resisted complying with Department of Education requests for data because the work required to produce those data fell outside the scope of the test contract.

The first error was made in computing the 1991 and 1992 anchor sets; the second involved the accidental destruction of records by CTB's scoring subcontractor, Data Recognition Corporation.

In several instances, Department of Education management exhibited poor judgment in attempting to address the issue of the nine percent decline in student performance on the 1992 reading test.

When the anomalous results of the form four anchor test became evident, department management should have requested more data and a more detailed explanation before proceeding with equating.

In dividing payment of Dr. Susan Phillips between two fiscal years, department management should have provided full disclosure of the fact that the payments were compensation for a single job and obtained appropriate approvals.

Although the revised version of Dr. Phillips' report was never used, department management should not have revised the document nor permitted CTB to rewrite the section on equating. Consultant's reports are advisory and the department is not obligated to accept consultants' advice. However, in this case, staff should have prepared a separate Department of Education report. That report should have openly acknowledged the existence of Dr. Phillips' report and explained the reasons why some of Dr. Phillips' recommendations were not acceptable.

Recommendations

Issue of 1992 Scoring

School districts will be provided all available information concerning results of the 1992 EWT reading test, including an explanation regarding the inconclusiveness of cutscore determinations. Each district will be provided the names of its students who fall within the two-point range of cutscores produced by various equating procedures. Districts will be advised to review each student's circumstances and to take whatever steps are needed, if any, to ensure fairness. The Department of Education will support districts' efforts in this regard.

Commitment to Statewide Assessment Program

The Department of Education strongly supports the statewide testing program and recommends that the various decision-making bodies of the state, including the State Board of Education, remain firmly committed to its continued implementation.

This testing program continues to be a critically important means by which the state: 1) pursues high educational standards for its system of public education; 2) assures the accountability of publicly funded educational institutions; 3) maintains the integrity of the high school diploma; and, most importantly; 4) assures that students receive a proper education.

The statewide testing program has effectively been achieving its intended purposes. Fifteen years ago, there was substantial evidence that significant numbers of our high school graduates lacked elementary-level knowledge and skills in communication and computation. The Minimum Basic Skills Test encouraged districts to rectify that problem. The HSPT established a higher standard by requiring for graduation ninth-grade knowledge and skills in reading, mathematics and writing. Districts focused their efforts on enabling students to meet this standard. Substantial improvements have been achieved.

The newest cycle in the state's testing program utilizes the HSPT-11 and the Eighth-Grade Early Warning Test to take an additional step toward raising performance standards in New Jersey. The testing program is part of a

broader effort to define important academic proficiencies, to teach those proficiencies, to measure students' acquisition of the proficiencies, and to use the knowledge gained from assessment in the continuous improvement of educational programs.

The HSPT-11 and the 8th-Grade EWT are sound, valid tests that provide the nucleous of the state's continuing commitment to improve the quality of New Jersey's public education system. Mistakes have been made in the overall administration of the testing program during its formative years. There is a sufficient amount of blame that can be assigned to various quarters of state government.

Yet, the choice that confronts New Jersey is that of either learning from past mistakes and enhancing its fundamentally sound quest for high standards, or of abandoning that quest prematurely in favor of a lesser effort.

The Department of Education recommends that New Jersey's commitment to pursue increasingly high standards and educational quality be maintained, and the statewide testing program be supported and improved.

Improving the Statewide Testing Program

The Department of Education recommends that the following steps be taken to improve the assessment program:

HSPT-11

Unlike the 8th-Grade EWT, the HSPT-11 will have been piloted for three years prior to its formal implementation as a graduation standard. An opportunity has been available to to work out problems of test design, administration and scoring. For graduation, students will have four opportunities to take the test, and those who do not pass will have access to alternate assessment procedures. In addition, the following steps should be taken:

- Implementation of the HSPT-11 for high school graduation in 1993-94 should go forward as planned, with the first administration being held in Fall 1993.

- The HSPT is a high-stakes examination. Therefore, results of open-ended questions should be excluded from the equating process unless and until there is specific and convincing evidence of the ability of scoring and equating technology to accommodate the effects of such items in a reasonably precise and predictable way.

- The Department of Education will aggressively seek resources to conduct formal validation studies within the next three months.

8th-Grade EWT

The Department of Education recommends that there be a recognition of the evolution that necessarily and appropriately occurred between 1991 and 1992

in the 8th-Grade Test. The educationally desirable, and predictable, result is a test that is better now, after two years experience, than that which was initially when developed under tight timelines with insufficient support.

The department further recommends that the 8th-Grade Test be administered as planned this month, and in subsequent years, with the following improvements:

- 1993 should be treated as a baseline year. Results for 1993 should not be equated with results from 1991 or 1992.

- Equating should be done for the first time when scores from 1994 are equated with scores from the new baseline year of 1993.

- In 1993, state standards on the 8th-Grade EWT should be represented by bands or ranges of performance rather than as a single cutscore.

- To identify future performance bands, equating should be conducted based upon embedded anchor items. A pool of 20 items should be drawn from the 1993 test to serve as embedded anchor items on the 1994 test. An additional pool of 20 items should be drawn from the 1994 base test to create a pool of 40 anchor items. In each future year, 20 new anchor items should be drawn and the 20 oldest should be eliminated.

- In 1994 and in subsequent years, open-ended questions should be excluded from the equating process unless and until there is specific and convincing evidence of the ability of equating technology to accommodate the effects of such items in a reasonable predictable and precise way.

- In 1994, results of the EWT should be pre-equated to 1993, under the supervision of an independent expert, before equating formula are formally applied to produce comparable performance bands.

- Development of the 1994 versions should include formal field tests validation studies.

- State monitoring code requires that 75 percent of each district's students achieve the state cutscore on the 8th-Grade EWT. This requirement is inconsistent both with the purposes of the EWT and the recommendations offered above. The State Board should reconsider this requirement.

- The Department of Education should intensify its efforts to educate local districts in the interpretation and use of 8th-Grade EWT results.

- The Department of Education should request waivers of the government employment freeze in order to properly staff the Bureau of Statewide Assessment. The vacant position of Director of the

Office of Educational Programs and Student Services, to which the assessment bureau reports, should be filled immediately. Staff positions vacated through layoffs should be refilled.

- Items on the EWT should be secured by having districts either return or destroy test booklets.

Commerically Produced Tests

The Department of Education does not recommend replacing the statewide testing program with commercially available tests:

- The quality of the state tests (both the HSPT-11 and the EWT) is not disputed. All experts, consultants and advisors praise their content and format.

- The New Jersey testing program is part of a broader state effort to identify essential academic proficiencies, to teach those proficiencies to students, to measure students' acquisition of the proficiencies and to improve educational programs based on analyses of results. Commerically produced tests are not designed to certify that New Jersey's students have learned core proficiencies approved by the State Board of Education. Rather, they are designed to compare individual students with other students in the nation against a sampling of commonly taught knowledge and skills.

- Commercially developed tests mainly have the potential to tell school districts how they compare with the national status quo, and they motivate districts to strive for that standard. In fact, because such tests are revised infrequently, they may sometimes encourage the pursuit of obsolete knowledge and skills.

- Commercially developed tests are not free of problems. The shared responsibility of two organizations (NJDOE and CTB) is one reason that problems of the 8th-Grade EWT were openly revealed and addressed. The department needs to exert greater control in its contracts with testing firms. If the department were to buy wholesale into a commercially produced test, it would have no control over test development, administration, or scoring.

- New Jersey has invested substantial public resources in the development and implementation of the HSPT-11 and the 8th-Grade EWT A premature and unwarranted abandonment of the program would waste this investment of public resources.

- Because commercially developed tests are infrequently revised, they are also not secure. Districts have an opportunity to learn the specific items contained on tests.

- Most states and federal agencies are moving away from norm-referenced tests and toward performance assessments. New Jersey would step backward if it were to abandon its efforts in favor of a return to norm-referenced tests.

LK/pc:1/1855f

BEST COPY AVAILABLE

REPORT ON THE EIGHTH-GRADE EARLY WARNING TEST (EWT)

Submitted to:

Mary Lee Fitzgerald
Commissioner of Education

By:

Thomas King
Director of Internal Audit

Leo Klagholz
Assistant Commissioner
Division of Academic Programs and Standards

New Jersey State Department of Education
225 West State Street
CN 500
Trenton, New Jersey 08625-0500

March 1993

12X

Report on Eighth-Grade Early Warning Test (EWT)

The Issue

In January 1993, staff of the Department of Education's Bureau of Statewide Assessment, informed Commissioner Fitzgerald and assistant commissioner-designee Leo Klagholz of a problem involving the Eighth-Grade Early Warning Test (EWT).

Bureau staff indicated that, in Spring 1992, the test contractor, CTB Macmillan-McGraw Hill (CTB), erred in its construction of a sample of students who took the anchor form of the reading section of the 8th-Grade EWT. The anchor form is that version of the reading test that was administered to groups of students who took two different versions of the base reading tests in 1991 and 1992. Results of the anchor test were used to establish a "cutscore" for the 1992 test that is comparable to the one used in 1991. This was accomplished through a statistical "equating" of students' scores for each year, using the performance of the sample of students who took the "anchor" test as a common denominator.

Staff of the Bureau of Statewide Assessment reported that in implementing this equating process, the contractor used a sample of 1200 "anchor-test" students that was not representative of the full group of students (approximately 33,000) who took the anchor test. Therefore, the statistical equating process had produced a cutoff score for 1992 that was two points higher than it should have have been. Staff indicated that, as a result, about 5,000 students had been identified as candidates for remediation who would not have been so identified had the proper cutoff score been used.

Staff reported that the problem was revealed through the following sequence of events:

- In reviewing test results in May or June 1992, assessment bureau staff noticed that about nine percent more students fell below the cutscore on the reading test in 1992 than fell below the 1991 cutscore. The decline was noticed after individual student scores had already been sent to school districts.

- A consultant was hired in July 1992 to work with staff of the Department of Education's Bureau of Statewide Assessment in identifying reasons for the apparent decline in student performance. Several possibilities were studied but no conclusive determinations were made.

- Staff of the department, assisted by CTB staff, met with district representatives in September 1992 to review the reading results, to discuss the findings of the consultant study, to explain the equating process, and to indicate a lack of information that would explain fully the drop in reading scores.

- In preparing the annual report on statewide results of the 8th-Grade EWT, Bureau staff found in October 1992 that the sample of student answer folders that was used by CTB to equate the reading section was not representative of the total group who took the test.

- Bureau staff asked CTB in October 1992 to conduct additional analyses to help determine the effects of the flawed samples, and CTB initially resisted complying with all aspects of the request.

- The issue was discussed in October 1992 with the Technical Advisory Committee (TAC), a group of nationally recognized measurement experts which serves in an advisory capacity to the New Jersey testing Program. The consensus of TAC members was that the equating procedures should be redone using the full population of 33,000 students who took the anchor test.

- In early December 1992, CTB provided the revised analyses which indicated that, if the total sample of students were used, the cutscore would be two points lower than that which was used in Spring 1992 to identify students who need remediation.

In order to provide immediate public awareness, Commissioner Fitzgerald briefly described the issue at the February 1993 meeting of the State Board of Education and announced the Department of Education's intention to examine the problem in the coming weeks.

CTB responded to accounts of this announcement by indicating that "CTB Macmillan-McGraw Hill made no mistakes in developing the statistical procedures for New Jersey's Early Warning Test or in scoring the test, which was administered to eighth-grade students in New Jersey in 1991 and 1992." CTB representatives indicated further that:

> The original analysis had been based on carefully selected representative samples of students. The reports issued in May were based on this analysis and showed that the percentage of students not meeting state standards was greater in 1992 than in 1991. As a consequence, somewhat more students were identified as candidates for remediation than were identified in 1991.
>
> The reanalysis based on the whole population resulted in a new reading standard that was lower than the standard used in the May score reports.
>
> The difference between results based on samples and results based on populations is to be expected. This concept is fundamental and well known to testing specialists. The differences do not imply that the sample results are faulty or wrong in any way.

Purpose and Scope of the Current Study

The current study was undertaken to resolve conflicting descriptions of the problem by developing a more comprehensive factual account of events. This factual account is intended to provide a basis for determining: 1) actions needed to resolve the immediate problem of the 1992 results; and 2) actions required to improve the state testing program and re-establish its credibility.

The current report is based upon information obtained from interviews with persons associated with the principal organizations, and a review of a substantial amount of documentation. The following persons participated in interviews:

Joel Bloom, former assistant commissioner, Division of General Academic Education, New Jersey Department of Education

Thomas Corcoran, former policy advisor for education, Office of the Governor

Richard DiPatri, former assistant commissioner, Division of Educational Programs, New Jersey Department of Education

Anne Fitzpatrick, research scientist, CTB Macmillan-McGraw Hill

Ellen Haley, director of contract management, CTB Macmillan-McGraw Hill

Michael Kean, vice president for public and governmental affairs, CTB Macmillan-McGraw Hill

Diane Kubinski, reading specialist, Bureau of Statewide Assessment, New Jersey Department of Education

Edward Masonis, director, Bureau of Statewide Assessment, New Jersey Department of Education

Eva Miller, data analyst, Bureau of Statewide Assessment, New Jersey Department of Education

Jason Millman, chairman, Technical Advisory Committee, New Jersey Statewide Assessment

Susan Phillips, professor, Michigan State University and consultant to New Jersey Department of Education

Cummings Piatt, former deputy commissioner, New Jersey Department of Education

Wendy Roberts, operations specialist, Bureau of Statewide Assessment, New Jersey Department of Education

Luis Salgado, former assistant commissioner, Division of Educational Programs and Student Services, New Jersey Department of Education

Calla Smorodin, HSPT 11th Grade Coordinator, Bureau of Statewide Assessment, New Jersey Department of Education

David Taggert, vice president of national accounts, CTB Macmillan-McGraw Hill

Jack Thompson, eastern regional evaluation consultant, CTB Macmillan-McGraw Hill

Wendy Yen, director of research and measurement, CTB Macmillan-McGraw Hill

It must be noted that the current study was conducted under severe time constraints. The equivalent of approximately seven working days during a

one-month period was alloted for interviewing, review and analysis of documents, and report writing.

This document should be considered an interim report of that information which could be obtained within the time permitted, and it should stimulate continued efforts to study and improve the state testing program.

New Jersey's Statewide Assessment Program

Individual districts, schools and educators are responsible for enabling each student to attain the highest possible levels of academic knowledge and skills, commensurate with the student's capabilities, interests and efforts. One role of state government is to support local initiative by providing funding, coordination, information, encouragement and other forms of assistance.

Another role of government is to represent the public interest, and particularly students' interests, by holding educational institutions accountable for providing all students the opportunity to reach at least a common minimum level of knowledge and skills. New Jersey's Statewide Assessment Program was established by law in 1976 in response to mounting evidence that this basic educational goal was not being adequately met. There were indications of basic skills deficiencies among substantial numbers of students.

The New Jersey Statewide Assessment Program serves both roles of state government. It is part of a broader effort to raise academic standards and improve educational programs. That effort involves the identification of important academic proficiencies, the development of educational programs that enable students to acquire these proficiencies, the evaluation of students' acquisition of the proficiencies, and the improvement of educational programs based on evaluation results.

The assessment program is also a public accountability mechanism. State government uses the results of the statewide tests to: 1) inform the taxpaying public of the relative performance of publicly funded educational institutions; 2) maintain a consistent definition of the high school diploma by assuring that each student is provided the essential knowledge and skills that the diploma represents; and 3) motivate appropriate state intervention in cases where institutions consistently fail to enable students to meet essential standards.

The Eleventh-Grade High School Proficiencies Test (HSPT-11) is the culmination of efforts to determine that level of knowledge and skills which represents a reasonable state-enforced minimum for high school graduation -- that standard which districts must be held accountable for enabling students to meet.

Beginning in 1981-82, all ninth-grade students were required to pass the Minimum Basic Skills test (MBS) as one requirement for the high school diploma. The MBS Test assessed elementary-level knowledge and skills in communication and computation. In order to raise the bar to a more acceptable level, the State Board of Education in 1983 approved the High School Proficiencies Test (HSPT) which requires ninth-grade knowledge and

skill levels in reading, mathematics and writing. Students were first required to pass the HSPT for high school graduation in 1985-86.

The HSPT-11 was established through legislation in 1988, and students must pass it as a requirement for the diploma beginning in 1993-94. The HSPT-11 sets the essential standard which the high school diploma represents at eleventh-grade proficiency in reading, mathematics and writing.

The 8th Grade EWT was established through the same legislation that instituted the HSPT-11. Starting with the year that began in September 1990, the EWT is administered annually to eighth-grade students to determine their progress toward mastery of state graduation proficiencies, so that those who are not making adequate progress can be provided appropriate assistance.

The focus of the current report is on the reading section of the Eighth-Grade Early Warning Test. The reading test presents the student with reading passages followed by questions concerning those passages. The test includes four types of passages: narrative, informational, persuasive and everyday. The questions include literal, inferential and beyond-the-text.

Development and March 1991 Administration of the Eighth-Grade EWT

The law instituting the HSPT-11 and the 8th-Grade EWT was enacted in December 1988 (see N.J.S.A. 18A:7C-6). The law requires that the HSPT-11 be administered annually for high school graduation starting in the school year that begins September 1993; it requires that an 8th-grade test be administered annually starting in the school year that begins September 1990.

During the 1988-89 school year, broad-based committees were formed to review curricula and identify skills to be assessed by the HSPT-11. These skills were reviewed by experts and shared for comment with all New Jersey school districts. The skills were finalized in June 1989.

During the 1989-90 school year, a similar process was used to identify skills for the 8th-Grade EWT, and to align the eighth-grade skills with those identified for the HSPT-11. In addition, committees were formed to develop test specifications, and the Department of Education formulated a budget request of $1,798,000 to support initial development and implementation of both tests.

In March 1990, the Governor requested that the Legislature approve the department's full recommended budget of $1,798,000 for test development and implementation.

On June 27, 1990, the Department of Education issued a Request for Proposals (RFP) to prospective bidders for the development and initial administration of the 8th-Grade EWT. At the time the RFP was issued, Saul Cooperman was Commissioner of Education, John Ellis was Commissioner-designee, Richard DiPatri was Assistant Commissioner of Educational Programs, Joel Bloom was Assistant Commissioner of General Academic Education and Stanley Rabinowitz was Director of Statewide Assessment.

BEST COPY AVAILABLE

The following characteristics of the RFP are relevant to the current study:

- The RFP required contractors to propose ways in which they would conduct the following validation studies of the EWT:
  - examination of test content by experts for domain representation and other content related issues;
  - study of whether test items are representative of the skill arrays and test specifications upon which they are based;
  - study comparing the content of the test with locally used achievement tests, the SAT and ACT, and locally constructed and/or district-wide tests.
  - study of dimensionality of the test;
  - study of the extent to which the K-12 curriculum covers the material on the test;
  - study of the impact of the test on school curriculum; and
  - investigations of test items for bias.
- The RFP indicated that "approximately 10-20 percent of all mathematics and reading items may be open-ended," and it required contractors to provide for scoring of open-ended questions should they be included.
- The RFP required contractors to field test essay questions. However, the current review found no provisions in the RFP for field testing multiple-choice or open-ended questions. Staff recalled that field testing of these items was excluded due to anticipated budget and time constraints, and that the test-design process compensated for this exclusion through preliminary item analysis. Preliminary item analysis is a statistical procedure for eliminating ineffective test questions from the scoring process after the test is administered but before it is scored.
- The RFP contained no provisions for equating 1992 results with those of 1991. Staff indicated that a determination had not yet been made regarding the use of a specific cutscore.

In a June 28, 1990 memorandum, Joel Bloom informed Commissioner-designee John Ellis of the desire of the Governor's office, with the concurrence of the department, to include open-ended questions on the HSPT-11 and the 8th-Grade EWT. Dr. Bloom stated that, "because of the substantial expense for the development and scoring of open-ended questions," only $550,000 would be left for the 8th-Grade EWT after funding of the HSPT-11. Dr. Bloom also stated that, as a result, an additional $350,000 would be needed to fund development of the 8th-Grade EWT. Dr. Bloom indicated further that Thomas Corcoran, policy advisor for education in the Governor's office, was seeking the additional funds.

In a July 27, 1990 memo to Commissioner Ellis, Deputy Commissioner Cummings Piatt and Assistant Commissioner Richard DiPatri restated the problem described in June by Joel Bloom. Drs. Piatt and DiPatri also indicated that the Governor's office had since determined that any additional funds for the 8th-Grade EWT would have to come from existing accounts of the Department of Education. Drs. DiPatri and Piatt outlined possible sources of funds, but recommended that a final decision be delayed until after contractor bids had been received.

Contractor bids were received on August 15, 1990, and the lowest bid ($1,079,332) was submitted by CTB. Because of interagency delays in the processing of bids, a decision to award the contract to CTB was not made until mid-October, five months before the 8th-Grade EWT would be administered to students.

During this period, there were numerous personnel changes in the Department of Education. Saul Cooperman resigned effective July 1, 1990. Joel Bloom left his position on August 13, 1990. Stanley Rabinowitz resigned effective September 7, 1990, Richard DiPatri left his position on September 14, 1990. Jacqueline Cusack, former assistant to Joel Bloom was named acting assistant commissioner of the Division of General Academic Education. Edward Masonis, formerly a staff member in the assessment bureau was appointed acting director of the Bureau of Statewide Assessment. Jacqueline Cusack was eventually laid off in March 1991, and Luis Salgado was subsequently appointed assistant commissioner.

Internal documents indicate that, in order to fund the development of the 8th-Grade EWT, the department amended its contract with CTB to eliminate the essay field-testing and validation studies indicated in the original RFP. This reduced the cost of the EWT contract by $55,850. In addition, the requirement that validation studies be conducted was removed from the contract for development of the HSPT-11. This released an additional $106,623 for development of the 8th-Grade EWT. The balance of funds needed for the EWT contract was transferred from other department accounts.

An internal written chronology of the Department of Education shows that representatives of CTB met with department staff in November 1990 to begin planning development of the EWT. CTB sent test items to the department in December 1990, and these items were reviewed by New Jersey test development committees in January 1991.

At this relatively late point in the test development process, several important determinations were made. First, it was decided, in response to pressures to use test results immediately for student placement and eventually for monitoring, that a specific cutscore would be established in the initial years of test administration.

Second it was determined that, if cutscores were to be set, then it would be necessary to equate the 1992 and 1991 cutscores.

Third, therefore, the EWT contract had to be renegotiated to fund the cost of establishing cutoff scores and of designing and implementing anchor materials for equating. Studies of test bias were also restored to the contract through these renegotiations.

Fourth, staff indicate that because of inadequate time and resources, anchor questions were not embedded in the base tests as is commonly done. Rather, a separate anchor test was created and divided into four separate subtests. Students selected to take the anchor test would have to do so after having taken the base test. Further, different students would take different anchor subtests.

In February 1991, a month before the test was to be administered, the New Jersey Association of School Administrators requested a one-year delay, citing a lack of information about the test and insufficient opportunities for district preparation. Internal records of the department show that staff, also concerned about the effects of tight timelines and inadequate resources, recommended consideration of several options: 1) postpone implementation of the EWT for one year; 2) postpone the initial administration until Fall 1991; 3) conduct the initial administration on a trial or "due-notice" basis; or 4) administer the test but interpret the results using "bands" of performance, rather than a single cutscore.

Following a conference phone call involving deputy commissioner Piatt, former assistant commissioner Bloom and director Masonis, a decision was made to proceed with test administration but to interpret results in the first year using performance bands.

March 1992 Administration of the 8th-Grade EWT

The 8th-Grade EWT was administered to students for the second time in March 1992. The test was scored, an equated cutscore was set, and results were sent to districts in June 1992.

On April 28, 1992, CTB research scientist Anne Fitzpatrick notified assessment bureau staff by memo of an anomaly in the performance of students who took the anchor test for reading. As noted above, the reading anchor test actually consisted of four separate subtests — one used narrative passages, one used informational passages, one used persuasive passages and one used everyday passages. The equating process included four samples of 1,200 students, each of which took one of the four forms of the anchor test — 4,800 students in all. No single group of students took the entire anchor test.

Dr. Fitzpatrick reported that "the sample that took the '92 Form 4 items [everyday text] were systematically different in some way from the samples that took the other forms." Specifically, Dr. Fitzpatrick reported further that, although the samples of students who took the 1992 forms 1-3 did slightly worse than their 1991 counterparts, the form 4 sample in 1992 did better than its 1991 counterpart. Dr. Fitzpatrick's memo offered no possible explanations for the anomaly; rather, it simply indicated her intention to exclude Form 4 from the equating process, stating:

> "Wild horses couldn't convince me to do otherwise. So there is no confusion about this remember that in theory we should be able to scale the '92 test using the items from one form alone, ... or from two forms, or some from one form, some from another, etc. So eliminating the five Form 4 items concerns me not the least; the equating will likely be as good....I don't know whether it is possible, but it might

be interesting to to try to figure out what (sic) the Form 4 group was different than the rest. Sometime when you have nothing else to do. Since the Form 4 samples did better than expected, look to see whether they have 'better' demographics than their counterparts on the other forms."

In separate interviews, assessment bureau staff members, Diane Kubinski and Eva Miller stated that they told Dr. Masonis of their concerns about Dr. Fitzpatrick's plan to eliminate Form 4 from the equating process. The staff members said that they recommended that CTB be required to conduct preliminary analyses to determine the effect that the removal of Form 4 might have on equating.

However, the department accepted Dr. Fitzpatrack's decision to remove Form 4 without requiring any additional analyses. Dr. Masonis stated in an interview that he accepted Dr. Fitzpatrick's decision in deference to her substantial expertise in the field of equating. He indicated further that no one disagreed with or advised against his doing so.

It should be noted that "pre-equating" procedures such as that Ms. Kubinski and Ms. Miller say they urged are, in fact, common practice. Test contractors take preliminary "dipstick" readings to determine the potential effects of equating and, where necessary, they make adjustments before equating formulae are applied. Assessment bureau staff indicated that pre-equating procedures were not used in scoring the 8th-Grade EWT because of a tight schedule and inadequate resources.

Performance Decline

In May 1992, the Department of Education received individual student scores from CTB. These scores were sent to local districts to be used in placing students in remedial courses the following September. When assessment bureau staff examined statewide results in more detail, they found that approximately 9 percent more students fell below the reading cutoff score in 1992 than in 1991.

Consultant Review

A nine 9 percent statewide decline in student performance on comparable tests is statistically possible but unlikely. Mainly for that reason, the Bureau of Statewide Assessment decided in May 1992 to employ a consultant, Dr. Susan Phillips of Michigan State University, to help determine reasons for the decline. At the same time, Eva Miller wrote to Anne Fitzpatrick requesting an explanation of the procedure that had been used to equate the 1992 scores. In a June 1, 1992 memo to Edward Masonis, Diane Kubinski referenced "problems with the statistics used and the equating study procedures." She wrote further that, "Anne Fitzpatrick indicates that the Form 4 was eliminated from the 1992 test results. This may or may not cause the results to differ from one year to another since Form 4 was not eliminated in 1991."

Dr. Phillips conducted her study during June and July and submitted a confidential draft report on July 24, 1992. This report describes many of the problems encountered and decisions made in the original design and

administration of the 8th-Grade EWT, some of which resulted from the limited amount of time and resources available for test development. However, the report concludes that these deficiencies were probably not the cause of the 9 percent decline in performance on the 1992 reading test.

Rather, Dr. Phillips concluded that the decline was more likely caused by problems in the equating process. In her report, Dr. Phillips pointed out that most equating procedures work best when the two tests to be equated are comparable or parallel. She cited the deletion of Form 4 from the 1992 equating transformation, the anchor test design, the inclusion of open-ended questions in equating, and other variables that might have influenced the comparability of the two base tests and, therefore, the precision of the equating procedures. However, when the consultant applied other equating models, both with and without Form 4, the results changed very little.

Department records indicate that Dr. Phillips was paid a total of $11,500 (plus expenses) at a rate of $500 per day for 23 days. The total amount of $11,500 exceeded the state maximum that can be paid without obtaining specific approval of the New Jersey Department of Treasury. Dr. Phillips was paid two separate amounts of $6,500 and $5,000 in two different fiscal years. The rate of $500 per day exceeded the normal state rate, and was approved on the basis of the department's indication that Dr. Phillips was as "sole source" vendor.

July 28 Meeting

A meeting was called on July 28, 1992 to discuss the implications of Dr. Phillips report. The meeting was attended by Anne Fitzpatrick, Susan Phillips, Jason Millman, Edward Masonis, Eva Miller, Wendy Roberts and Diane Kubinski.

In advance of the meeting, Susan Phillips assisted assessment bureau staff in preparing ten questions concerning the equating process. These questions were shared in advance with Anne Fitzpatrick, who was asked to submit written answers that would guide discussion at the July 28 meeting. Dr. Fitzpatrick declined to answer those questions that required researching the test data, stating, "CTB does not have a data tape now, nor is it scheduled to receive one until mid-August."

Participants characterize the July 28 meeting as a professional discussion between Ann Fitzpatrick and Susan Phillips, mediated by Jason Millman. In essence, Dr. Fitzpatrick argued that the equating procedures used by CTB are theoretically sound; Dr. Phillips asserted that, although the procedures are sound theoretically, they may not have worked in the specific practical application of the 8th-Grade EWT.

Assessment bureau staff indicate that most participants in the July 28 meeting felt that Dr. Phillips recitation of problems that occurred in test design were not relevant because: 1) the effect of the identified problems on test quality was to some degree a matter of conjecture; and 2) Dr. Phillips' herself dismissed their relevance to the decline in student performance on the reading section of the test.

Most participants in the July 28 meeting indicate that there was general agreement that the Department of Education would request that CTB provide data needed for a more thorough analysis of the equating issue.

Supervision By Upper Management

At this point, it should be noted that neither Commissioner Ellis nor Assistant Commissioner Luis Salgado was involved in attempting to resolve the problem of the EWT. Dr. Masonis indicated that both were briefed on the matter in June, and he recalled that Commissioner Ellis reacted angrily to information about the nine percent decline in student performance on the reading test. Dr. Masonis said that Commissioner Ellis made a statement to the effect that: this is an election year, and I don't want any bad news. Other assessment bureau staff recall Dr. Masonis having recounted the commissioner's comments to them at the time.

However, all staff members strongly assert that their actions were motivated solely by their desire for accurate test results. At the same time, staff members felt that their superordinates were inattentive to the reading issue and to the testing program generally. They believed that they were left to fend for themselves and expected to work out difficult problems without generating controversy. File records support staff members' contentions that they kept their supervisors informed, but received little direction or support.

Requests for Data

On August 3, 1992, Dr. Masonis wrote to to Dr. Fitzpatrick requesting the additional data that participants in the July 28 meeting agreed was needed. Dr. Fitzpatrick provided only some of the requested data. Assessment bureau staff indicate that some of the data were provided on disks that were formatted in a way that made them difficult to use. Dr. Fitzpatrick attributed the problem to a lack of reformatting expertise on the part of assessment bureau staff. Dr. Fitzpatrick also provided a written response to the observations that Dr. Phillips made in her draft report concerning the effects of several variables on equating. In essence, Dr. Fitzpatrick indicated that while some of the observations may be factually accurate, their effects on the equating process is unknown.

District Workshops

In September 1992, a decision was made to conduct a series of workshops for school district representatives to provide information about: 1) the development, scoring and equating of the EWT; 2) the quality control procedures used to verify test results; and 3) the Department of Education's study of the reading results. Department staff were joined by Anne Fitzpatrick in conducting the workshops at three regional locations: Cherry Hill, New Brunswick, and Wayne. All districts were invited to send representatives to these meetings. The issue of the reading results was discussed openly with district representatives, who were reminded that the EWT results are but one basis for deciding which students should be assigned to remediation.

BEST COPY AVAILABLE

This latter point is one which the department has consistently stressed in its communications with school districts. For example, in April 1991, Luis Salgado wrote to all superintendents concerning the 8th-Grade EWT saying:

> Score standards will be provided to help districts assess the performance of their students on the EWT. The distribution of all possible scores on the EWT will be divided into several bands. Students who score in the lowest band do not meet state standards.
>
> Students whose scores fall in the lowest band do not automatically have to be assigned to basic skills instruction in grade 9, however. The EWT scores are to serve as the primary indicator, but the final decision about assigning students to basic skills instruction must be based on EWT scores and additional assessment information about the student...The basic skills instruction is intended to supplement rather than supplant regular coursework (N.J.A.C. 6:8-6.3), so EWT scores should not be used to determine academic tracking.

This same information was reiterated in the regional meetings that were conducted in September 1992.

Revision of Consultant's Report

Prior to the September 1992 meetings, staff of the department's Bureau of Statewide Assessment revised Susan Phillips' report on the EWT reading results. Dr. Phillips was called by Edward Masonis, and by Eva Miller, and asked whether she would convert her report, which was still a draft, to final form in such a way that would reflect the consensus decisions made at the July 28 meeting. Dr. Phillips' recalls being asked to revise the draft report in a way that would provide a "political slant."

Dr. Phillips indicates that she refused to make the requested revisions as a matter of principle because she saw her task as being to provide a straightforward analysis of the issues as she saw them, and because she felt that her draft report accomplished that task. Dr. Phillips said that she told Department of Education staff that they could use her report in any way they desired as long as her name was not associated with any other document than the one she submitted.

Eva Miller was assigned the task of revising Dr. Phillips' draft report and, in an interview, stated that she resisted this assignment. File documents show that Edward Masonis deleted numerous passages from those sections of the draft report that primarily analyzed the design of the EWT. Eva Miller incorporated Dr. Masonis' modifications into her revision of the consultant report. In addition, Anne Fitzpatrick of CTB was permitted to rewrite those sections of Dr. Phillips' draft report that addressed equating. The cover page of the final report represents the document as one of the Department of Education, and Susan Phillips' name is not indicated.

The resulting document is one which duplicates most of Dr. Phillips' report verbatim, but with numerous additions and deletions. Dr. Phillips did not receive a copy of the final document. Despite having granted the department permission to use her draft as it saw fit, Dr. Phillips said in an interview that she did not expect that revisions would be made to what otherwise is a verbatim copy of her draft report.

Dr. Masonis indicated that the second version of the report was prepared because the department had decided after discussions with CTB and TAC members, not to accept all of Dr. Phillips recommendations. Dr. Masonis indicated that the second version was prepared to reflect the department's final position. The latter document was to be used in response to possible future inquiries concerning the rationale for the department's actions.

Assessment bureau staff either destroyed their copies of Dr. Phillips' original draft or returned them to Dr. Masonis at his direction. However, at least one copy was retained. There are no indications that the department's version of the report was ever used in any way. When Dr. Masonis was asked to provide a copy of the consultant report he provided both the Phillips' draft and the department's version.

Representativeness of Samples

In early October 1992, staff of the assessment bureau began analyzing statewide results on the EWT in preparation for their annual report to the State Board of Education. Staff indicate that, in so doing, they discovered that the sample of students used to equate the reading results was not representative of the full population that took the reading anchor test -- that is, the sample did not perform at the same levels as the full population.

Therefore, correspondence was exchanged between Edward Masonis and Anne Fitzpatrick concerning Dr. Masonis' request for additional data on the samples and the equating process. Some of the data were provided.

In an October 8, 1992 memo to Dr. Masonis regarding the calibration sample, Dr. Fitzpatrick stated:

> I did make a computational error in computing the differences in the 1991 and 1992 anchor sets, although the correct results still imply that the 1992 population was somewhat less able than the 1991 population. I can't figure out how I made the error since I checked my results twice; I suspect that the error resulted from my doing the calculation by hand. I'm glad you asked me to check the computations once again before you published them...I apologize for the error.

In a memo dated October 12, Dr. Fitzpatrick wrote to Dr. Masonis stating, "I have been told by the Director of Research and Measurement, Wendy Yen, that the work you are currently requesting is beyond that covered by the contract. If I am to do any further analyses, further discussion about how the work will be paid for and when it can be done will have to occur before I can do any more."

Request for Re-equating

On October 19, 1992 Dr. Masonis wrote to three members of the TAC -- Drs. Millman, Hambleton and Brown -- explaining the assessment bureau's finding that "the total score on the reading anchor was different for the calibration sample when compared to all students who took the anchor...for both 1991 and 1992." The issue was discussed with the three TAC members in a conference call on October 21.

As a consequence of that discussion, Dr. Masonis wrote the next day (October 22) to David Taggert, CTB's vice president for customer products, formally requesting that CTB:

> Equate the 1992 and 1991 EWT Reading test sections using the entire sample of students who took the anchor forms in 1992 (a total of 33,095 students) and the calibration sample for each of the forms in 1991 (about 1,200 students per form).
>
> Equate the 1992 and 1991 EWT Reading test sections using the entire sample of students who took each of the anchor forms in both 1992 and 1991. The total sample who took the 1991 Reading anchors is 36,258 students.

In response, CTB indicated that the additional work would cost $11,440.

On October 29-30, 1992, the TAC met to discuss the issue and recommended that: 1) the reading results for 1992 are not likely to be an accurate reflection of an actual drop in achievement; 2) the report to the State Board should be postponed until further investigations are done; and 3) calibrations should be rerun using the entire sample of students who took each set of Reading anchor items.

On November 2, 1992, Dr. Masonis authorized CTB to proceed with the understanding that the cost of $11,440 would be paid.

On November 10, 1992, Anne Fitzpatrick indicated that she would proceed with the first of the equating runs that Dr. Masonis requested on October 22 (see above). However, she recalled that Dr. Masonis's second request had been determined by the TAC to be "unnecessary because it is pointless." Dr. Masonis responded to Susan Woodward of CTB that Dr. Fitzpatrick's recollection was inaccurate.

Finally, on December 11, 1992, CTB produced equating results that were based on the full population of students who took the anchor forms of the EWT. When Form 4 is excluded from these population data, the results are not substantially different from those obtained in May 1992 using samples of students. When Form 4 is included in the population data, the equating process produces a cutoff score that is two points lower than that originally indicated.

Assessment bureau staff assert that this latter procedure is the correct one because: 1) it includes all available data; and 2) it produces a result -- no decline in student performance -- that is more probable than those produced by other formulae. CTB maintains that the procedure using the population data is different from, rather than necessarily superior to, the procedure based on student samples. In an interview, CTB officials stated that there is a sufficient amount of error built into double-equating procedures to account for the differences in the results they produced in this case.

Destruction of Answer Sheets

It should be noted that, during the course of the current study, two New Jersey parents requested copies of their children's answer sheets. In response, CTB's scoring subcontractor, Data Recognition Corporation (DRC) of Minnetonka, Minnesota, indicated that New Jersey answer sheets had inadvertently been destroyed. In a February 10, 1993 letter to CTB vice-president, David Taggert, Malgana Hallstrom, DRC's director educational services wrote:

> DRC had only the demographic and multiple-choice portions of the EWT answer documents. The constructed-response and writing task portions were microfilmed and then returned to the schools. DRC understood that we were to store the answer documents for one year following test administration. After one year we would request written approval from NJDOE to destroy the documents. After written approval was received from the client, written authorization would be given to DRC's warehouse to destroy the documents...In the case of the 1992 EWT documents, there was a miunderstanding when the warehouse supervisor thought he received verbal approval to destroy the documents.

The destruction of original answer sheets violates the state's contract with CTB. DRC maintains computerized records of answer sheets, which can be reproduced electronically.

HSPT-11

Finally, the focus of this inquiry was exclusively on the 8th-Grade Early Warning Test. However, it should be noted that due to additional budget cuts, validation studies for the HSPT-11 still have not been initiated. In addition, it should be noted that the assessment bureau office has lost four staff positions to layoffs at a time when time when the new monitoring code requires expansion of the testing program under short timelines.

LK/pc:1/1853f

March 8, 1993

The Honorable John Rocco
Chairman, Assembly Education Committee
CN 068
Trenton, NJ 08625

Dear Mr. Chairman:

I regret that I was unable to accept your invitation to appear before the Assembly Education Committee on March 15. The contracted research in which I am currently engaged requires me to be out of state conducting interviews from March 14 through March 19. I appreciate your invitation to appear at the hearing, and I would welcome an opportunity to do so at another time.

Since some recent press reports have stated that I directed the Department of Education to include open-ended items on the Early Warning Test (EWT), and since I cannot attend Monday's hearing to respond to these allegations, I am sending you this letter to let the members of your Committee know what role I played in the development of the EWT. **I want them to understand that to my knowledge no one in the Governor's Office issued any specific directions about the testing program in June or July of 1990 or at anytime while I served as Policy Advisor to Governor Florio.**

Before I elaborate on the events that occurred in the spring and summer of 1990, I want to make my views about assessment policies in New Jersey clear. While I have always supported the state assessment program, and believe that it plays a critical role in raising performance, I also believe that we need high quality tests to assess higher order skills and the application of knowledge in the academic disciplines. We must develop assessment strategies which encourage the adoption of challenging curriculum and ambitious teaching. We cannot continue to assess only the basic skills, and expect to challenge our students to work towards higher levels of attainment.

I believe that current efforts to focus more attention on important concepts and skills in mathematics, science and social studies will flounder unless these disciplines are adequately covered by a high-quality assessment program. If we are serious about wanting our schools to adopt more challenging curricula and wanting our young people to work harder to reach higher standards, then we must decide what our students should know and be able to do and develop curricula and assessments consistent with those goals and standards. This means that we must learn to use open-ended items, and other approaches to assessment.

Other states are developing curricular frameworks consistent with the emerging

national standards in the disciplines and strategies for assessing student mastery of this more challenging content, including performance tests that assess problem-solving skills and application of knowledge. New Jersey should be an active participant in this national effort to raise academic standards and develop better means of assessment.

I will now turn to the current controversy surrounding the development of the EWT. It is alleged in the Department's report on the EWT that in early 1990 someone in the Governor's Office expressed a desire to have open-ended questions added to the EWT and that the Department's efforts to comply with this desire created a budget problem for the testing program. This, in turn, led to the Department's decision to cancel validation studies for the EWT and the HSPT. The implication is that this chain of events led to the problems recently encountered with the EWT.

Last week, after the press reports alleging my involvement in these events, I spoke to Dr. Bloom. He told me that there were documents in the Department indicating the agency's intent to include open-ended items on the EWT that predated any discussions he and I had had in the spring of 1990. **Therefore, the suggestion in the Department's report that the inclusion of open-ended items resulted from a desire on the part of someone in the Governor's Office is simply incorrect.** The fact is that the desire to include such items emanated from within the Department at an earlier time.

Joel Bloom and I did meet several times during the spring of 1990 to discuss the development of the Governor's Program for Excellence in Mathematics and Science and the GoodStarts program. Typically we met in Dr. Piatt's office, but we also met separately on at least two occasions - May 11 and June 26. These latter two meetings were after the development of the RFP for the 11th grade test which included open-ended items. At some point, Dr. Bloom must have briefed me on the development of the EWT, and told me about the Department's plans to include some open-ended items on the tests. I certainly would have supported such an initiative. **However, I do not recall giving any directives to include open-ended items, nor do I recall discussing how many such items would be on the test.**

Language is important in this case. In the memo from Dr. Bloom to Dr. Ellis cited in the Department's report on the EWT, it refers to "the desire from the Governor's Office, with our concurrence based upon improved technology, to include open-ended "performance" measures on the tests." There is no indication of how such a desire was expressed, by whom, or when. A statement such as "I wish we were doing more of this" is an expression of a desire. But it is not a directive or an order. Furthermore, the expression of such a desire on my part or anyone else's would not relieve the Department's staff and their contractors from their responsibilities for making sound decisions about the test specifications and their obligation to ensure that the resources to support the test were adequate, the work done technically, and the tests valid.

I am certain that I issued no directives concerning the items on the test because it

would have been totally inconsistent with my relationship with the Department at that time. I was quite formal in my dealings with the Department and all of my communications were channeled through Dr. Piatt. That is, no actions were taken even on routine information requests without his knowledge, and usually his active involvement. I simply would not have directed someone in the Department to take an action such as altering the EWT. If I had wanted to make such a recommendation, I would have raised the issue with Dr. Piatt, and he would have arranged a discussion with other senior members of the Department. I recall no such discussion about the EWT. I held Dr. Piatt, Dr. DiPatri, Dr. Bloom, and other senior members of the Department in great esteem, and I would not have jeopardized my relationship with them by issuing a directive which I had no authority to issue.

Moreover, I played no role in subsequent decisions affecting the EWT made during the administration of Dr. Bloom's successors, Dr. Cusack and Dr. Salgado. Nor did I have any discussions with Dr. Rabinowitz or Dr. Masonis, Directors of the State Assessment Program about the EWT. If I had taken the initiative to to direct someone to add items to the test, I think that I would have followed up to see that it was done and how it worked out. **There was no such follow-up, because I had no involvement in the decisions made about the EWT.**

I want to thank you for this opportunity to clarify the record and reiterate my willingness to speak before the Committee at another time. I hope that you and the other members of the Committee will not let the problems experienced with the EWT undermine your support for the state assessment program and that you will support Commissioner Fitzgerald's efforts to correct the problems and move forward.

I further urge you and your committee to examine the new approaches to assessment being undertaken in Connecticut, Kentucky, California, Vermont and other states, and to consider the possibilities that they offer for the improvement of curriculum and pedagogy. The New Jersey state tests are important tools for accountability, but we also should be concerned about the ways in which they affect standards, curricula, teaching, and tracking. New Jersey would benefit from hearings at which teachers, parents, and national experts were invited to give testimony about what our children need to learn and be able to do, the kind of teaching that we want for our children, and the implications of these aspirations for state and local assessment programs. I urge you to take the leadership in promoting such a statewide dialogue.

Sincerely,

Thomas B. Corcoran

RECEIVED
MAR 1 0 1993
BY

Route 1, Box 141
Wimberley, Texas 78676

March 6, 1993

The Honorable John A. Rocco
Chairman, Assembly Education Committee
295 West Route 70
Cherry Hill, N.J. 08002

Dear Chairman Rocco:

Reporters have telephoned me about a report by Leo Klagholtz that attributes to me not releasing 8th grade early warning test data, apparently for political reasons. Such an assertion is egregiously false.

Reporters have also stated that you are holding a hearing on this alleged "coverup." I'm glad you are reviewing this matter, but I suspect you will be confronted with the usual conflicts and political rhetoric.

If a report critical of my actions were to be written, common sense, as well as basic fairness, would suggest that I be asked for my perspective about the events. Obviously, that was not done which creates grave suspicions in my mind about the author's motives.

I have not seen the report. My sole knowledge about it comes from reporters who have called me to inquire about my response to some of its statements.They quote various issues and ask for my reaction. Based on this limited information let me respond to the salient issues as I understand them.

1. At no time was I aware of any attempt by anyone in the department to withhold test data information. If that were the case, I had no personal knowledge of it.

2. The possible drop in test scores was not the issue. Test scores will vary. Some years go up. Some go down. This is normal. The trend over time is the key. The major concern was to insure that the test scores were valid so they could form a solid basis for analyzing how to improve student learning. Reporting a drop or an increase in scores based on flawed data would have been misleading. The integrity of the system was at stake.

3. My clear recollection is that Mr. Ed Masonis, director of testing, and Luis Salgado, former assistant commissioner who supervised this testing, shared their concerns about the preliminary results for the second year of the 8th grade early warning test. They said:

   a. They were uncertain that the results were accurate.

   b. There were anomalies in the results that caused them to worry that some assumptions were made in the scoring by the testing company that were in error.

   c. They were working with the test company to assess the issues more completely and determine if the results were accurate.

BEST COPY AVAILABLE

d. While some additional time would be required to complete the process thoroughly, the staff could not assure me that the preliminary results were reliable.

It is eminently clear that no one is well served if the results are inaccurate. No matter how painstaking it may be, the State is entitled to solid, accurate data that are free from statistical juggling whether intentional or unintentional.

I encouraged the staff to work with the company to insure accuracy. In a previous test (from a different company) an error occurred in a diagram that didn't affect the results of the test but the media had a field day with the error. The department would understandably be subject to severe criticism if test data conclusions were released that proved to be flawed. The delay was not political, it was sensible.

Yes, there was a shortage of department staff due to budget constraints and yes, there is a complex developmental process that will require improvements and refinements as experience is acquired with the new assessments. The entire curriculum and testing process will take a decade to put in place on a world class basis. Anyone who represents that this is simple or can be done quickly without trial, error and revision, or by merely giving attention to certain policy matters creates a serious distortion. We like to find demons to blame or quick solutions rather than buckle down to the serious work at hand.

Reporters say the report states that I had warned the staff it was an election year and we didn't want any bad news as the reason why the results weren't released. That is inaccurate. My key concern was with the accuracy and integrity of the data. I said so. I advised that we had a previous error; the department's testing budget had been reduced by the legislature and was at risk; that it was an election year, which is "crazy season" at best, and would expose the department to a political fusillade if our final results were inaccurate; that New Jersey's leadership in working on higher order thinking skills would be imperiled if we could not maintain integrity in the data; and that students would ultimately be shortchanged if we did not maintain a challenging, effective system of assessment.

No one can assert with accuracy that I did anything other than strive for integrity and honesty in reporting. That is my trademark and, while it might seem unrealistic in a state as political as New Jersey, it is a fact.

I hope the raising of the testing issue will enable everyone to revisit some of the basic issues and insure that New Jersey has a first rate system of testing that improves instruction for students. New Jersey has an enviable record of leadership, but far more should be done and resources are scarce to do the tasks well. This is why I strongly recommended to the State Board of Education that New Jersey become a member of the New Standards Project so New Jersey could have access to some of the best testing minds in the country at an expense New Jersey could afford.

New Jersey's earlier state tests were laudable but too easy. They did not require the challenge our students need to be competitive in a complex world marketplace. Measures to promote higher order thinking skills are currently being developed, but have not yet reached the degree of reliability and validity we will need. Some problems along the way are inevitable. But don't let problems deter you or conclude that little leadership has been given to the task. That is disingenuous.

Recall from history the race between the horse drawn carriage and Tom Thumb, that upstart steam engine. Yes, the belt slipped on the steam engine and the horse had one last victory. But the future should have been clear. So it is with testing. We are on the verge of some valuable assessment processes that will challenge students to learn. There is an adage that "What is tested will be taught---what is not tested will be ignored." While one can argue that such an adage is not always operative, there is sufficient truth in it to raise the special concern that tests should focus on the crucial concepts and skills we want students to learn. That includes higher order skills and the demonstrated ability to perform.

There is one more issue I know you raise, and I respect your concern for it. You ask, "Why not simply buy tests 'off the shelf' that are available from commercial tests makers and forget New Jersey's efforts to build its own tests?" This would save money and be simpler, the argument goes. It is a responsible question that implies a reasonable alternative.

There are several reasons why New Jersey's course is particularly promising. First, commercial tests are improving, but they still lag behind the best instructional practices. New Jersey would settle for second best if it would abandon an aggressive attempt to be a leader in developing improved measures.

Secondly, it is extremely difficult to keep tests confidential when every school district buys its own tests and keeps them on hand. This is one of the reasons a leading test critic decries the standardized test results that show "everyone in the nation is above average." Commercial tests purchased by individual districts are subject to this national "Lake Wobegon" syndrome. Test security is difficult to maintain in the best of circumstances and, candidly, the state has had problems with some school districts even with our strict security.

Finally, it helps instructional practice and student learning---the reasons for giving the tests---when the professionals in the state are involved in learning about testing and when they help to set the standards and participate in forming the assessment system. There is greater ownership and growth. This process needs to be improved not abandoned.

Some commercial tests are good and will be useful in New Jersey especially for specific content areas. But don't let an "off the shelf" test, which tends to have a lowest common denominator appeal, set the standards for graduation in New Jersey. Our students deserve better.

Let me summarize. The issue of coverup is phony. The issues of how to develop the most challenging tests that will help students improve are real. Much hard work lies ahead. Get on with it. Don't be diverted by the oldest game in the book: Blame one's predecessor. As they say in Texas, "That dog won't hunt."

Good luck to you in your deliberations. I hope you will separate the wheat from the chaff of polemical writers, that you will focus on the best practices, and that you will eschew the temptation to score political points at the expense of public understanding. Legislative leadership in education is crucial to progress in the state and I wish you every success in your efforts to focus on what is best for students.

Sincerely,

John Ellis

xc: Media

BEST COPY AVAILABLE

13 Genesee Ct
Bayonne, NJ 07002
March 1, 1993

Assemblyman J. A Racco
295 First Route 70
Cherry Hill, NJ 08002

Assemblywoman M. V. Weber
Center of Doubletree, N. Delsea Dr.
Glassboro, NJ 08028-1400

Dear Assemblywoman Weber and
Assemblyman Racco:

I am the parent of a child who I believe has been caught up in the 1992 E.W.T. problem. Having graduated 8th grade with High Honors and a 92 average in reading, she was informed she failed the reading section of the E.W.T by 2 points. Her teacher and Principal spoke at length with the Comprehensive Education Director, insisting an error had been made. Their efforts were in vain, and my daughter was placed

34X

BEST COPY AVAILABLE

in a Remedial Reading Class for one period every day in the 9th Grade at Tacoma High School. Her Skills teacher has commented to her recently that she does not belong in the Class.

My complaint now is that the computed test-scores have been available since December, but have not been released to the high schools. I spoke to the director of the program at high school & she informed me that everything is on "Hold" as per instructions from the State Dept. of Education. However, schedules are now being planned for next year and without the Corrected scores available Guidance Counselors must go under the Assumption that a child will still be in the Remedial Class next year. I feel this is unfair.

If it is found a child passed the E.W.T. and will not be in remedial Classes next year, why not let them know now, so they

BEST COPY AVAILABLE

can plan their schedules accordingly and pick their electives.

I realize the entire testing program is under investigation, as it well should be, but releasing the corrected scores would not affect the investigation. Now that it's been learned that the State Dept. of Education made a big mistake, they should admit it and correct their mistakes. Let the records straight for the children's sake, instead of covering up your mistakes. Also, I feel that any child wrongly placed in a remedial class should have that removed from their school records. The State made the mistake, why should the children pay for it, which is exactly what the State Dept. of Ed. planned to do when they tried to bury their mistake.

Sincerely,

Mrs. Dixie Harmon

cc: Robert J. Braun
Star Ledger

36X

BEST COPY AVAILABLE

MEASURING UP

Achieving World-Class Standards in Mathematics Education

Mathematics leads the nation in establishing standards for education

CURRICULUM STANDARDS
What students should know

TEACHING STANDARDS
How students should be taught

ASSESSMENT STANDARDS
How to determine what students are learning

Mathematics is front page news

It is the stuff of business and science, of politics and sports. Without mathematics, the world is a mystery. It contributes to scheduling, planning, and investing, as well as to the improvement of aircraft, heart valves, and software.

Mathematics is more than arithmetic

It is estimating and thinking, inventing and communicating. It is searching for patterns and solving problems. Mathematics is nourishment for 21st century minds.

But for lack of mathematics, our children are starving

Their future — and ours — depends on the mathematics they learn, yet most of them learn very little.

• • • • •

Our children CAN be better off mathematically

We know that every child can learn mathematics.

We know how to teach mathematics better than we do.

We know the standards we need to attain.

Our children WILL be better off mathematically

If **NATIONAL LEADERS** insist on standards-based systemic reform
. . . rather than on isolated programs of little lasting value.

If **STATE LEADERS** demand measurement of what's worth learning
. . . rather than just what's easy to measure.

If **THE PUBLIC** becomes convinced that perseverance is the key to learning
. . . rather than inherent ability or a "genius gene."

If **EDUCATIONAL LEADERS** promote local implementation of national standards
. . . rather than tolerating the status quo.

If **ADMINISTRATORS** provide continuing standards-based staff development
. . . rather than imposing demands without proper support.

If **SCHOOLS** expect all students to pursue a common core of mathematics
. . . rather than shunting many students, primarily minorities, onto dead-end tracks.

If **MATHEMATICS TEACHERS** engage students in active, minds-on tasks
. . . rather than in repetitive, mind-numbing exercises.

If **ALL STUDENTS** study mathematics each year they are in school
. . . rather than being turned off and dropping out.

• • • • •

Today's mathematics is for everyone

In the world of work, "figuring out" is replacing "figuring." It is time for America's leaders to help our children measure up.

The Mathematical Sciences Education Board

MISSION

To improve mathematics education for all students at all levels

Systemic Change . . . *Standards-based and system-wide*
National Leadership . . . *Forging consensus and stimulating reform*
Sustained Investment . . . *Building effective support structures*

Students, parents, teachers, administrators, policy leaders and the public

40 million elementary school children
. . . *and 2 million teachers*
25 million secondary school mathematics students
. . . *and 200,000 teachers*
8 million college and university mathematics students
. . . *and 50,000 teachers*

National education goals supported by local implementation

Guided by national goals
. . . *Supporting the NCTM Standards*
Cognizant of local autonomy
. . . *Respecting regional variation*
Motivated by national needs
. . . *Promoting social and economic well-being*

New Reports and Policy Statements

1989 *Everybody Counts: A Report to the Nation on the Future of Mathematics Education*
1990 *Making Mathematics Work for Minorities: Framework for a National Action Plan*
1990 *Reshaping School Mathematics: A Philosophy and Framework for Curriculum*
1990 *On The Shoulders of Giants: New Approaches to Numeracy*
1991 *Counting on You: Actions Supporting Mathematics Teaching Standards*
1991 *For Good Measure: Principles and Goals for Mathematics Assessment*
1991 *Moving Beyond Myths: Revitalizing Undergraduate Mathematics*
1993 *Measuring Up: Prototypes for Mathematics Assessment*

(Published by the National Academy Press)

1989 *Curriculum and Evaluation Standards for School Mathematics*
1991 *Professional Standards for Teaching Mathematics*

(Published by the National Council of Teachers of Mathematics)

1989 *Reshaping Undergraduate Mathematics*
1991 *A Call for Change: The Mathematical Preparation of Teachers of Mathematics*
1992 *Heeding the Call for Change: Suggestions for Curricular Action*

(Published by the Mathematical Association of America)

THE MATHEMATICAL SCIENCES EDUCATION BOARD

was established in 1985 by the National Research Council, which is the principal operating agency of the National Academy of Sciences and the National Academy of Engineering. The mission of the MSEB is to provide a continuing national capability to assess the status and quality of education in the mathematical scien[illegible]

MSEB

For further information, please contact:

MATHEMATICAL SCIENCES EDUCATION BOARD
National Research Council
HA 476
2101 Constitution Avenue, NW
Washington, DC 20418

Telephone: 202-334-3294
Fax: 202-334-1453
E-mail: mseb@nas.edu

BEST COPY AVAILABLE

38x

STATEMENT OF PRINCIPLES ON SCHOOL REFORM IN MATHEMATICS AND SCIENCE

The U. S. Department of Education and the National Science Foundation agree that all children should receive a challenging education in mathematics and science based on world-class standards beginning in kindergarten and continuing every year through grade 12. We therefore declare that we will act in concert to improve the teaching and learning of mathematics and science in the United States in order to advance the nation towards the fourth National Education Goal. In our collaboration, we will adhere to the following principles:

GENERAL

Communities, states and the federal government must work in a collaborative partnership to improve mathematics and science education.

The educational enterprise is a system with many parts, all of which must change in concert to meet the requirements of the 21st century.

NATIONAL STANDARDS

National content standards for students (what children should know and be able to do) must be developed and utilized as the basis for all other improvement activities, including instructional practices, assessment, and teacher preparation.

National content, assessment, and teacher preparation standards will serve as the foundation for grants to states to reform curriculum frameworks and local curricula, and for reform of instructional methods, textbooks, teacher education and certification, inservice programs, and student assessment.

CURRICULA

States should develop comprehensive standards-based K–12 curriculum frameworks, which establish a sequential program of learning in mathematics and science for all children.

The use of new technologies and their influence on increasing student achievement in mathematics and science should be supported through research and development activities at national centers, regional laboratories, and other pertinent institutions.

Textbook publishers and developers of instructional materials should ensure a change in their products to support the new national content standards through improved instructional practices such as problem-solving activities, creative student learning tasks and cooperative learning.

Curricula should promote active learning, inquiry, problem solving, cooperative learning, and other instructional methods that motivate students.

TEACHER EDUCATION AND CERTIFICATION

States should ensure that teacher education prepares new teachers to teach all children in accordance with the new national student content standards and the new state curriculum frameworks.

States should change teacher certification so that only highly and appropriately qualified and well-prepared persons, fully familiar with the content standards, requisite teaching practices, and improved assessment of knowledge are accepted into the profession of teaching.

States should adopt means of recertifying current teachers to ensure that all elementary and high school math and science teachers understand the national content standards and new instructional methods in mathematics and science.

Institutions of higher education, states, and local school districts should ensure that the preparation of new teachers is a joint responsibility of university faculty in arts and sciences and education in collaboration with school practitioners and departments of education.

Student Assessment

For purposes of accountability, states should develop new student assessments based on national content standards and state curriculum frameworks. These new assessments should test students' knowledge and understanding of mathematics and science in ways that are more complex and demanding than current tests.

Conclusion

In conclusion, we urge every parent, every school, school district, and state to insist:

On higher content standards for all students studying mathematics and science from kindergarten through high school;

On teacher preparation, inservice, and certification programs supporting the standards;

On a challenging K–12 curriculum that not only informs our children but inspires their understanding and enjoyment of the wonders and power of science and mathematics;

On the inclusion of all children, and particularly those who have been historically under-represented in a challenging curriculum every year;

And on fair and appropriate assessment instruments to measure student, school, and state progress toward this most challenging national education goal.

For Further Information

National Science Foundation
(202) 357–9522

U.S. Department of Education
Office of Elementary and Secondary Education
(202) 401–0657
Office of Educational Research and Improvement
(202) 219–2164

December 1992

PIP 93–1500
EDOERI 93–5

BEST COPY AVAILABLE

Statement of Principles

An agreement by the U.S. Department of Education and the National Science Foundation

TESTIMONY BEFORE NJ ASSEMBLY EDUCATION COMMITTEE – APRIL 19, 1993

Joseph G. Rosenstein
Director, New Jersey Mathematics Coalition

Mr. Chairman, committee members, let me introduce myself. My name is Joseph G. Rosenstein and I am in the Mathematics Department of Rutgers University, New Brunswick. I am here today as Director of the New Jersey Mathematics Coalition, which is a partnership of the education, business, public policy, and public sectors of New Jersey, all working together to improve the mathematics education of our children.

I am here to tell you that those who are actively involved in improving mathematics education in the state strongly support the directions taken in the mathematics portion of the High School Proficiency Test (HSPT) and the Early Warning Test (EWT). I will take a few moments to explain.

The basic fact, whether we like it or not, is that teachers teach to "the test", and that districts write curriculum based on "the test". That makes it particularly important that we test what we really value, that the tasks that we give children on our assessments reflect the tasks that we want them to perform.

For a number of years, standardized tests have focused on asking multiple choice questions which measure rote learning – and schools have responded by focusing on rote learning. It is no wonder that many students are turned off to mathematics.

There is almost universal agreement today that rote learning is not enough for citizens of the next century, that it will not enable us to achieve President Bush's goal of being first in the world in mathematics and science by the year 2000. That goal may be too ambitious, but there is no doubt that we can do much better – and there is a national consensus on how to improve.

As I indicated in an article published recently in the Star Ledger's Education Forum (copy attached), Colorado Governor Roy Romer gives the following simple advice "You have to know where you're going before you figure out how you're going to get there!" The direction endorsed by former Secretary of Education Lamar Alexander, was that of "standards-based education".

"Standards-based education" requires us to develop a vision of what we value, articulate that vision in clear statements of what we want to accomplish (called "standards"), and then figure out how we get there. With a goal and a plan in place, we will be able to assess how we're doing and take corrective action as necessary.

The goals for mathematics education were spelled out in a number of national reports over the past few years – including the *Standards* of the National Council of Teachers of Mathematics. These reports recommend that we focus more of our attention on developing in our children reasoning and problem-solving skills. What we value is that our children learn to reason about situations which they will face in their studies and in their careers, as consumers and as citizens, that they learn how to formulate problems and solve problems which arise in the real world.

BEST COPY AVAILABLE

Improving mathematics education has three components – content, assessment, and instruction. After determining what we value – as described in the national reports, we must next devise assessments which reflect what we value, and then we must ensure that our instruction addresses both what we value and what we assess. Are there assessments which reflect what we value? I must tell you that the HSPT is a large step in the right direction.

I will give a few examples. A standard problem is to give a rectangular shape and ask students to find its perimeter. Now any student knows that if a problem involves three or more numbers, the only way of solving it is add up all the numbers. And indeed, research shows that many students who have no understanding of perimeter get the right answer. Contrast this with an open-ended problem on the HSPT which asks students to build several complicated shapes out of simpler ones, such as four congruent squares, and find their areas and perimeters.

Or contrast a simple addition problem on standardized tests with one on the HSPT which provides a calorie table and asks students which of several meals come in at fewer than 800 calories.

Another open-ended problem on the HSPT involves a fast-food manager who has to use information about hourly gross receipts to schedule her employees so that their time is not wasted. Still another involves a driver who has to determine whether he needs to stop for gas before reaching his destination, and to estimate whether he will arrive on time for a job interview.

Solving each of these problems requires a deeper understanding of the mathematical concepts and more reasoning and problem solving than is traditionally required of students. I urge you to look at the mathematics portion of the HSPT, and to ask yourself if you value this kind of problem-solving; unfortunately, you will not find it – yet – on standardized tests.

We are setting high standards because we believe that all of our children can achieve more. Parents in other countries believe uniformly that the key to success is mathematics is persistence and hard work; only in this country do many believe that the key to success in mathematics is innate talent. The commitment of the mathematics community is to all students, as exemplified by the "Math 4 All" logo of the New Jersey Mathematics Coalition.

The most effective (and perhaps least expensive) way of communicating to districts, schools, and teachers what we value is by a statewide assessment which communicates our standards of mathematics education and insists that all students can achieve these standards. With the proper support, our teachers will ensure that our children will indeed meet those standards.

In a letter to Commissioner Fitzgerald (copy attached) in February, a copy of which was sent to Chairman Rocco for distribution to this committee, we indicated our belief that "this type of statewide assessment program offers a strong incentive and mandate for improvement of mathematics education across the state." We encourage the committee to support this program.

Section One: Page 53
THE SUNDAY STAR-LEDGER, February 28, 1993

Success in math, science requires high standards

The writer is a professor of mathematics at Rutgers University and director of the New Jersey Mathematics Coalition.

By JOSEPH G. ROSENSTEIN

"You have to know where you're going before you figure out how you're going to get there!" This simple advice was the basic thrust of Colorado Gov. Roy Romer's keynote address to the Feb. 9-11 education conference sponsored by the National Science Foundation on the theme "Beyond National Standards and Goals: Excellence in Mathematics and Science Education K-16."

Although the context of Romer's remarks was mathematics and science education, his advice applies, of course, to education in general. In discussing our educational system as a whole, or the curriculum of an individual school, we have at present no good way of evaluating how we're doing, since as a society we have no clear idea of what we value.

"Standards-based education" requires us to develop a vision of what we value, articulate that vision in clear statements of what we want to accomplish (called "standards"), and then figure out how we get there. With a goal and a plan in place, we will be able to assess how we're doing and take corrective action as necessary.

Why are standards needed, and why now? This question was addressed in a companion keynote address by economist Ray Marshall, formerly secretary of labor and presently professor at the Lyndon B. Johnson School of Public Affairs of the University of Texas.

The economic success of the United States in the first half of this century was based to a large extent on mass production, as exemplified

EDUCATION FORUM

by the automobile industry. Assembly-line workers did not need to know how to think in order to do their jobs – that was left to the managers; at best, they needed to be literate in order to read their instructions. Our educational model has been a two-tier system, with the "Three R's" for the many and thinking for an elite.

That model, Marshall argues, doesn't work anymore because of the emergence of technology, which renders obsolete many mindless occupations, and the globalization of the economy, which forces us to compete in unfamiliar areas.

In response to this challenge to be competitive, we may either increase productivity or decrease wages; Marshall notes that by failing to develop a strategic plan to increase productivity we have "backed into" the lower-wages solution, the results of which are increasingly evident.

The key to increased productivity is quality education, since the investment which yields the greatest return is investment in human resources, a theme echoed by President Clinton. And quality education requires a commitment to high standards.

Standards-based education has now become a major direction in education as the result of the success of the "Curriculum and Evaluation Standards for School Mathematics" developed by the National Council of Teachers of Mathematics (NCTM) between 1985 and 1989. While the implementation of the NCTM standards has only just begun, the document has clearly served as a rallying point (another meaning of "standard") for those seeking to improve mathematics education.

The response to the NCTM Standards has been very positive. It was endorsed by the National Governors Association, now chaired by Romer and previously by Gov. Bill Clinton, and was supported by former Secretary of Education Lamar Alexander, former governor of Tennessee.

A major project is now under way to develop science standards by the end of 1994 (under the leadership of the National Academy of Sciences), and other content areas are not far behind.

Quality education based on high standards must be modeled to

From preceding page

all students. The NCTM Standards asserts that all students can learn and do mathematics. In other countries, children are successful with mathematics; our children can enjoy the same success.

In a recent study, the students at the best of 20 Minneapolis schools were on a par with those at the worst of 40 schools in Taiwan and Japan. Parents in other countries believe uniformly that the key to learning mathematics is persistence and hard work; only in this country do many believe that the key to success in mathematics is innate talent. The commitment of the mathematics community is to all students, as exemplified by the "Math 4 All" logo of the New Jersey Mathematics Coalition.

Developing high standards for all students, world-class standards which will enable us to compete with world-class economies, is a major challenge for those committed to education.

In New Jersey, the Department of Education has established panels which are wor to articulate a vision and dev content standards in eight d ent areas.

In addition, the New Je Mathematics Coalition, toge with the Department of Ed tion, has obtained a three- grant from the U.S. Departr of Education to permit u translate the "where do we to go?" of the standards int "how do we get there?" tha follow. The project involve veloping a mathematics cu uum framework and wor with schools on implemer the framework.

Readers are invited to s with the coalition their sponses to "Where do we wa go?" More concretely, we your responses to the follo question: "What do we want high school graduates to k and be able to do in mathe ics and science so that they succeed in careers and in coi as citizens and as consume Let us hear from you! Our dress is NJMC, P.O. Box 1 New Brunswick, 08906.

42 X 117

BEST COPY AVAILABLE

New Jersey Mathematics Coalition

P.O. Box 10867, New Brunswick, New Jersey 08906

908/932-4065 *** FAX: 908/932-3477
Email: joer@math.rutgers.edu; patsy@dimacs.rutgers.edu
Rutgers University Center for Mathematics, Science, and Computer Education

February 24, 1993

Mary Lee Fitzgerald
Commissioner of Education
225 West State Street
CN 500
Trenton, New Jersey 08625

Dear Dr. Fitzgerald,

In light of recent media coverage of state testing programs, the New Jersey Mathematics Coalition would like to take this opportunity to affirm its support of the concepts and principles underlying the present testing program in mathematics (EWT and HSPT11).

Since teachers often "teach to the test", it is important that our assessment instruments reflect what we value, the standards agreed upon by the community. In recent years, the New Jersey assessments in mathematics have moved closer to reflecting the recommendations of the nationally recognized *Curriculum and Evaluation Standards for School Mathematics* of the National Council of Teachers of Mathematics. This is a direction which we applaud. At the Board of Education meeting in August 1992 we strongly endorsed the use of calculators on statewide assessments, and view the Board's resolution as potentially having a great benefit for instruction in mathematics.

We believe that this type of statewide assessment program offers a strong incentive and mandate for improvement in mathematics instruction across the state. We support the Department of Education in its efforts to retain and strengthen the program, and we stand ready to assist the Department in this effort.

Yours truly,

James W. Fella
Hoechst Celanese
Chair, Board of Governors

Janet Caldwell
Rowan College
Chair, Public Policy Committee

Joseph G. Rosenstein
Rutgers University
Director

cc: Anne Dillman (for distribution to Board of Education)
John Ewing (for distribution to Senate Education Committee)
John Rocco (for distribution to Assembly Education Committee)

43X

Testimony Before Assemblyman John Rocco
April 18, 1993
Patricia Clark Kenschaft, Ph.D.
Professor of Mathematics, Montclair State and Director of PRIMES
Project for Resourceful Instruction of Mathematics in the Elementary School

It is widely acknowledged that New Jersey's schools merit drastic change. As one who has worked with a dozen districts in the past five years, including Newark, Paterson and Passaic and several of the richest districts too, I can attest to the fact that dramatic improvement is needed and possible everywhere, but especially in the districts where children have least.

It is a sad truth that in modern America, education follows testing. Teachers nod when a speaker asserts that all education ceases for two weeks preceding standardized testing. Children are taught the useless, anti-intellectual skills needed in "scoring high."

My conclusion is that we need fewer standardized tests, and we need a radical change in those we give. If we look beyond the borders of our own country, we can build on the experience of others, since ours is the only country that uses standardized multiple choice tests. However, we can't just copy tests elsewhere because we live in a different culture.

Radical change requires experimentation, and experimentation almost always includes mistakes. We need more honest and humane acknowledgement that mistakes are inevitable everywhere. All children make mistakes. All faculty at all levels of education make mistakes. All administrators, guidance counselors, test makers, legislators, and newspaper writers make mistakes.

It is important that we not make rules that are so rigid that the mistakes of a few harm many. In particular guidance counselors, parents, and teachers need to be able to individualize the use of test results. Children should be placed using a variety of criteria, as I believe has been the official policy. Tests should not be the sole criterion. On the other hand, they must have some consequences or they will not be taken seriously by those taking the tests.

However, no testing system, no matter how wise, can itself improve education. Only helping teachers can do that.

I must take this opportunity to raise another issue that I believe is crucial to the improvement of New Jersey's educational system. We now have no statewide requirements in mathematics for teachers in the primary grades.

I have found teachers extremely eager to learn mathematics, but our state provides almost no support in their doing so. Furthermore, the educational establishment, both at the school and collegiate level, acts as if mathematics were expendable in both teacher preparation and in-servicing. Most educational leaders themselves have not had a thorough collegiate experience in mathematics.

Consequently, many New Jersey primary grade teachers, through no fault of their own, do not have a sufficient understanding of basic concepts such as areas, fractions, and subtraction to be able to teach them with the confidence and competence that little children need. By the age of ten our children are so damaged that I cannot repair the damage, although I can easily teach these concepts to younger children.

Therefore, our state has developed an expensive and wasteful remedial system that attempts to undo the damage we have inflicted on our children before the age of ten. Questioning this large remedial system may

#4X

jeopardize many jobs in an unstable economy, but almost all the people who would be displaced are capable of helping K-4 teachers and probably would find it far more satisfying than struggling with damaged, discouraged youngsters.

I hereby appeal to you, Mr. Rocco, to consider what can be done to help little children mathematically before they are intellectually and psychologically damaged. After five years of working with teachers, I can assure you that they are clamoring for help, help that apparently can come only from dynamic state-level leadership.

In summary, I support:

1. The administration of vastly fewer standardized tests. We need to save time and money for real education.
2. Developing tests other than multiple choice tests. Yes, they will be more expensive, but the current system of testing is very damaging to our entire educational enterprise, which is very expensive and absolutely central to the survival of our society and species.
3. Being more patient with each other and more tolerant of mistakes by everyone.
4. More flexibility in how test scores are used.
5. Most important:

A dramatic, state-wide initiative to help teachers in the primary grades learn mathematics so that they can teach it right the first time. Let us take steps to change our culture from one that punishes to one that collaborates in helping people do their jobs to the very best of their ability. In five years of circulating around many schools of New Jersey, I am strongly impressed with how much teachers want to help children and how capable they are of learning if taught. Let us help them help children!

45X

TESTIMONY

Education Committee Hearing
on Testing, April 19, 1993 Room 8
Trenton, New Jersey

My name is Dr Ira Sweet and I have been employed as a teacher , Guidance Counselor and school Psychologist in New Jersey Public Schools for over twenty seven years . I have also been associated with Trenton State College, Kean College and Ocean County Community College as an adjunct facilty in Psychology for over twenty years.

I appreciate the opportunity given to me to testify at this hearing and it is a sad affair that I have to be here. When I marched with Martin Luther King in Montgomery Alabama over 25 years ago, I was there to demonstrate for social justice and against institutionalized racism and discrimination. It is now twenty five years later and I must now speak out against institutionalized descrimination which exists within the STate Department of Education as it relates to testing procedures and practices.

Understanding the nature of these hearings, I should let you know beforehand that my remarks are related to testing problems in the public schools in New Jersey but not directly related to the test currently being investigated. You are concerned about 5000 students being placed in remedial programs because of inappropriate testing. I am concerned about students being classified as mentally retarded and placed in inappropriate special classes and programs because of inappropriate tests being used. I will let you be the judge as to which problem is the most damaging to children.

I am here to ask you to expand the scope of your investigation to find out why personnel from the State Department of Education are

46X

knowingly allowing school districts to use an intelligence test that is not valid, that discriminates against handicapped children, that is culturally biased and that is racially discriminatory.

Let me make you aware of the N J Administrative Code, Title 6 Chapter 28, titled Special Education, which became effective in July 1992. This is a revision of an earlier code. You should understand that this Code was adoped from guidelines set down by the U S Office of Education. In order for the State to get millions of dollors each year from the U S Goverment, they must agree to certain non-discriminatory guidlines .

Under 6:28-2.5 entitled " Protection in Evalative Procedures" it indicates that tests used to evaluate, classify and place handicapped students should be:

1. validated for the purpose for which they are administrated
2. That they take into consideration the pupil's cultural background, and language abilities
3. Take into consideration the pupil's handicap.
4. That the test not be racially or culturally biased

Again under 6:28-3.5 it states that the tests used should be reliable, valid and normed on a representitive population.

You should understand that when parents agree to have their children tested, they sign a form. This is a contract between them and the school district that the administrative code will not be violated. As of this time, this code is being violated and it is a betrayal to parents and their children. The laws of New Jersey and the federal goverment are simply being violated

Let me be more specific and identify for you a particular Intelligence test that is being used in one or more school districts in in New Jersey that does not meet the above mentioned requirements.

It is called the Slosson I Q Test. I would like to quote from this test's manual:

> Some persons have lived in what could be considered a barren, impoverished , deprived environment, they might not have attended school, not learned to read and may not have listened to radio or television --- The I Q obtained on persons from impoverished environments must be interpreted with much caution.

The manual then goes on to expand upon why the test should be used with much caution with individuals with the following problems:

1. Persons who are handicapped
2. Persons who have reading problems
3. Persons who have language problems
4. Persons who are brain damaged
5. Persons who are emotionally disturbed

Despite all of the precausions just cited, this test is still being used in one or more school districts in New Jersey for the purpose of identifying handicapped children and placing them in special classes and programs.

Let me cite from the Seventh Mental Measurements Yearbook. This is probably the ultimate authority on all published tests that are used in this country. It states:

The Slosson I Q Test appears to be a quick screening device ----- -----. However, the uncritical use of the Slosson I Q Test as a substatute for the Wechsler or Binet instruments is ill advised

It goes on to say:

> Extreme caution should be taken in relying on the Slosson I Q Test scores in situations where important diagnostic decisions are required, such as special class placement etc etc. The heavy emphasis on language skills makes it a difficult test for children who, for cultural and individual reasons, have language problems.

Gentlemenn, this is exactly was is continuing to happen.

Even if you question the opinions of the those who wrote their opinions in the Mental Measurements Yearbook, I wish to assure you that if you brought a thousand books related to tests and measurements and brought it into this room, you would find the same opinions that I have cited today.

All that I have presented is not new. As far back as 1979 litigation between Larry P and the Superintendent of Public INstruction of the State of California concerned itself with minority children being over placed in classes for the mentally retarded based upon the results of invalid I Q Tests. The federal district court determined that there were violations of:

Title VII of the Civil Rights Law of 1964
Education of all Handicapped Children Act of 1974
14 Ammendement - Equal Protection of the Law

One of the results of the litigation was the demand made by the federal goverment that all states receiving federal funds for special education include in their codes a section regarding protection in evaluation procedures as I have previously cited.

Let me briefly cite from my doctoral diseration dated 1976:

This writer reviewed the statistical data from the State DEpartment of EAucation dated 1973 on children classified and placed in special education programs in New Jersey. An analysis of the reports indicated that 56.1 percent of the children in mentally retarded classes were white and 43.9 per cent were from minority groups (Blacks, Puerto Rician, Cubans). These findings confirm those reported by Mercer which indicated that a disproporttionate number of children in mentally retarded classes were from minority groups.

In 1985 the results of a study by the Division of Special EDucation was published in the newspapers. Their findings were similar to mine.

More recently, on March 14, 1993 there was an article in the N Y Times entitled: " Educators seek Panel to keep Tests Bias-FRee" Let me read you sections from this article,

A group of leading educators is calling for the creation of an independent commission to make sure that national student tests are free of sexual, racial and cultural bias.
A growing body of research suggests that stedent performance can be dragged down by unintended bias in the wording of test questions.

BEST COPY AVAILABLE

As a child advocate, on October 5, 1992, I filed a complaint of non-compliance with the New Jersey Division of Special Education. My complaint was related to evaluation procedures in that they were not in compliance with special education laws. Along with my complaint, I submitted copies of psychological evaluations that showed that students had average intelligence with valid I Q Tests and that the same children had I Q scores in the mentally retarded range three years later with the Slosson I Q Test.

Shortly thereafter, I was amazed to receive a letter from the Division of Special Education stating " The Division of Special Education has reviewed your complaint and has determined that your concerns are not within the jurisdiction of the Division to investigate in accordance with N J A C 6:28-9.2. This statement of the Division of Special Education is in complete contradiction of N J A C 6:28-2.5 entitled " Protection in the Evaluation Procedures" which I have previously cited.

Even if the N J A Code was not present, it is surely the public policy of this State not to discriminate against handicapped children, not to discriminate against minorities, not to allow for biased testing and not to act in ways that promotes racial discrimination.

Gentlemen, there are some questions that requires answers ! Why is the State Department of Education going against public policy? Why does it choose to ignore the rules and regulations of the Administrative Code ? Why is biased testing being permitted to be used in the schools of New Jersey ? If the Division of

BEST COPY AVAILABLE

Special Education does not have the jurisdiction to investigate complaints concerning discriminatory testing, then who has the responsibility to investigate ?

I don't know the answers to these questions. Perhaps you can provide me with the answers.

Again, I I want to thank you for allowing me to testify. If have any questions, I would be most happy to answer them.

Ira Sweet
1 Roselle Court
Lakewood, N J
908-364-1232

BEST COPY AVAILABLE

8 Lahiere Ave.
Edison, N.J. 08817
April 25, 1993

Alan J. Steinberg
Assembly Majority Office
State House, CN-098
Trenton, N.J. 08625-0098

Dear Mr. Steinberg;

As you requested in our phone conversation on April 21, 1993, for purposes of expediency, I am attaching portions relevant to my testimony of the "Standards for " Standards for Educational and Psychological Testing," prepared by the Committee to Develop Standards for Educational and Psychological Testing of the American Educational Research Association,The American Psychological Association, and the National Council on Measurement in Education, 1985. For purposes of brevity, I will use "Test Standards," and cite the page and standard for documention in my testimony. I am enclosing copies of some of my articles as part of the evidence I am presenting against the current state's student testing program before the Assembly Education Committee. I have reasons to believe that this testing program should be abolished.

I am a life long resident of the state and a tax payer. I have had extensive experience in state and county

facilities working with children and families as a Clinical Psychologist. My experience included psychological assessment and testing, staff development and training, supervision of psychological interns mainly in the area of psychological and educational testing, and treatment of children and families.The problems treated included child abuse which involved the preparation of evaluations for the court and appearing as an expert witness on matters pertaining to children. My articles on topics including advocacy, psychological assessment, professional child abuse, neuropsychology, rehabilitation,and attitudes of professionals toward clients have appeared in numerous journals.I have also presented workshops and been a panelist at many national and regional conventions.

I understand that there is reason to suspect that the "Test Standards" were not followed regarding the development of the tests (p.25, 3.1; p.26, 3.3-3.5; p.28, 3.16; p.29, 3.21, 3.22). Chapter 8, p. 49 refers to "Educational Testing and Psychological Testing in the Schools," and practically all of the standards listed in that chapter are pertinent to violations of the "Test Standards" by the state Department of Education.

I also understand that the Department of Education developed and administered the tests used in the skills testing program without having standardized or validated them before giving the tests to students throughout the

state. This is a violation of the "Test Standards" p.13, 1.1, 1.2; p.20, 2.1. 2.2. 2.3..

I seriously question the predictive value of the current testing program. For example if a child is ill or in a bad mood, the child would not do as well as when in a good mood or physically well and rested. A child not motivated for testing may do poorly one day, and when motivated, do better on the same test another day. It is well known that some individuals are poor test takers and traumatized by testing which apparently was not taken into account by the state Department of Education which administers mass testing.

There is evidence that the tests used in the current testing program apparently discriminate against minorities, ethnic groups, and the disabled. The testing program has not demonstrated that it is culture free. Disabled students may not, because of their disability, be able to take timed tests. Many students do not understand English. I agree with the concerns expressed in Assemblyman Garcia's testimony and the legislator who charged that standardized tests are a form of "institutional racism." All of the "Test Standards" in Chapters 13, pp. 73-75 and 14, pp. 77-80 are relevant for present purposes and reflect violations of the "Test Standards."

All standardizded tests are based on norms which entail the comparison of each child with other children in a group.

BEST COPY AVAILABLE

Such a comparison is a violation of the uniqueness of each individual and discourages the inherent creativity of all children. If a child does not meet the test norms, then the interpretation is often made that there is something "wrong" with the child-the child is labeled ab-normal and therefore does not "fit in" with the established norms. These norms are often not relevant to the educational, cultural, or socio-economic status of a particular child, so that children, are, in the final analysis, compared on the basis of erroneous normative data. I consider all standardized, mass testing a waste of time and money, abusive to children, and they have nothing to do with the measurement of actual or future attainment in real life situations.

In her testimony, Mrs. Oppenheimer told of the problems she had when she tried to obtain information about her son's performance when he scored below the cut-off score for reading. The problems Mrs. Oppenheimer encountered could be considered a violation of the "Test Standards" p.84, 15.10, p.85, 16.2; p.86 16.4, 16.5..

Test scores often have an impact in determining the course of a child's scholastic and vocational life, and could even deprive children of a livlihood of their choice. I have reasons to believe that the state's current testing program is a form of professional child abuse. It is not cost effective and I fully support the testimony of Assembly Minority Leader Joseph Doria. As a tax payer, I resent this

large expenditure of money which could be used for direct services to children, especially when nationally recognized, standardized educational tests are available in abundance.

If mass testing is an absolute requirement, I would strongly recommend as Assemblyman Doria asdvised in his testimony, that the State Department of Education use nationally available tests. I would add, using only those tests which fully meet the "Test Standards." For purposes of brevity, my other recommendations are indicated in the enclosed articles which can be directly applied, in principle, to the state Department of Education's testing program.

I am sorry that I was unable to be present at the Assembly Education Committee Hearing on April 19th. If I can be of any further assistance or there are any questions regarding my testimony, please feel free to contact me.

Sincerely yours,

Louise Mead Riscalla, Ph.D.

Louise Mead Riscalla, Ph.D.

without regard to the group from which a person comes. Differing regression slopes or intercepts are taken to indicate that a test is differentially predictive for the groups at hand.

Under these circumstances, a given predictor score yields different criterion predictions for people in different groups and a given criterion score yields a different predictor cut score for people in different groups. If fitting the common prediction equation for the two groups combined suggests that the criterion performance of people in either group is systematically overpredicted or underpredicted, one possibility is to generate a separate algorithm (e.g., regression) for each group. Another possibility is to seek predictor variables that reduce differential prediction without reducing substantially overall predictive accuracy. If separate regressions are considered, the effect of this decision on the distributions of predicted criterion measures for the two groups is usually of interest.

Several proposed ways of evaluating selection bias rest on different definitions of the fairness of a selection procedure. Unlike selection bias, however, fairness is not a technical psychometric term; it is subject to different definitions in different social and political circumstances. At present a consensus of technical experts supports only one approach to selection bias as technically appropriate. This approach is adopted in the *Standards* with the understanding that it does not resolve the larger issue of fairness.

A quite different usage of the term differential prediction arises in the context of placement or classification. In that context evidence is needed to judge the suitability of using a test for classifying or assigning a person to one job versus another or to one treatment versus another. It is possible for tests to be highly predictive of performance for different education programs or jobs without providing the information necessary to make a comparative judgment of the efficacy of assignment or treatment.

Standard 1.1

Evidence of validity should be presented for the major types of inferences for which the use of a test is recommended. A rationale should be provided to support the particular mix of evidence presented for the intended uses. *(Primary)*

Comment:
Whether one or more kinds of validity evidence are appropriate is a function of the particular question being asked and of the context and extent of previous evidence.

Standard 1.2

If validity for some common interpretation has not been investigated, that fact should be made clear, and potential users should be cautioned about making such interpretations. Statements about validity should refer to the validity of particular interpretations or of particular types of decisions. *(Primary)*

Comment:
It is incorrect to use the unqualified phrase "the validity of the test." No test is valid for all purposes or in all situations. If a test is likely to be used incorrectly for certain kinds of decisions, specific warnings against such use should be given. On the other hand, no two situations are ever identical, so some generalization by the user is always necessary. Test developers should present their validation evidence in a way that can aid such generalization.

From: "Standards for Educational and Psychological Testing," American Psychological Assn., 1985, reprinted 1990

57X

BEST COPY AVAILABLE

which typically varies more from one group of test takers to another, is less informative.

Estimates of the reliability of a test should consider not only the relevant sources of error, but also the types of decisions anticipated to be based on the test scores and on their expected levels of aggregation (individual versus groups of test takers). For example, tests are sometimes used as the primary basis for making dichotomous decisions. In testing to determine certification for successful completion of a course of study, the primary interest is in the decision. Of course, there may be more than two categories, but the pass-fail or mastery-nonmastery decision is common. Estimates of the consistency of decisions are needed whenever decision rules assign people to categories according to specified test score intervals. An estimate of the standard error of measurement at the cut score is helpful.

Standard 2.4

Standard 2.1

For each total score, subscore, or combination of scores that is reported, estimates of relevant reliabilities and standard errors of measurement should be provided in adequate detail to enable the test user to judge whether scores are sufficiently accurate for the intended use of the test. ***(Primary)***

Standard 2.5

Comment:
It is not sufficient to report estimates of reliabilities and standard errors of measurement for only total scores when subscores are also reported. The form-to-form consistency of total scores on an achievement test may be acceptably high, for example, yet content specific subscores may still show relatively poor form-to-form consistency. When subscores are reported, users need information about their reliabilities and standard errors of measurement.

Composite scores may take various forms, some of which may be expected to be more reliable than the parts and some less reliable than the parts. But scores representing differences between scores obtained from two tests or from repeated administrations of the same test (called *gain scores*) are generally less reliable than either of the parts. Scholastic ability tests are often used to predict achievement test scores, and the difference (or residual score) is computed between the observed score and the predicted score. Such residual scores are, in general, less reliable than the original observed scores. Another composite score that is usually less reliable than are its parts is a ratio score.

Standard 2.2

The procedures that are used to obtain samples of individuals, groups, or observations for the purpose of estimating reliabilities and standard errors of measurement, as well as the nature of the populations involved, should be described. The numbers of individuals in each sample that are used to obtain the estimates, score means, and standard deviations should also be reported. ***(Primary)***

Standard 2.6

Standard 2.7

Standard 2.3

Each method of estimating a reliability that is reported should be defined clearly and expressed in terms of variance components, correlation coefficients, standard errors of measurement, percentages of correct decisions, or equivalent statistics. The conditions under which the reliability estimate was obtained and the situations to which it may be applicable should also be explained clearly. ***(Primary)***

BEST COPY AVAILABLE

which typically varies more from one group of test takers to another, is less informative.

Estimates of the reliability of a test should consider not only the relevant sources of error, but also the types of decisions anticipated to be based on the test scores and on their expected levels of aggregation (individual versus groups of test takers). For example, tests are sometimes used as the primary basis for making dichotomous decisions. In testing to determine certification for successful completion of a course of study, the primary interest is in the decision. Of course, there may be more than two categories, but the pass-fail or mastery-nonmastery decision is common. Estimates of the consistency of decisions are needed whenever decision rules assign people to categories according to specified test score intervals. An estimate of the standard error of measurement at the cut score is helpful.

Standard 2.4

Standard 2.1

For each total score, subscore, or combination of scores that is reported, estimates of relevant reliabilities and standard errors of measurement should be provided in adequate detail to enable the test user to judge whether scores are sufficiently accurate for the intended use of the test. ***(Primary)***

Standard 2.5

Comment:
It is not sufficient to report estimates of reliabilities and standard errors of measurement for only total scores when subscores are also reported. The form-to-form consistency of total scores on an achievement test may be acceptably high, for example, yet content specific subscores may still show relatively poor form-to-form consistency. When subscores are reported, users need information about their reliabilities and standard errors of measurement.

Composite scores may take various forms, some of which may be expected to be more reliable than the parts and some less reliable than the parts. But scores representing differences between scores obtained from two tests or from repeated administrations of the same test (called *gain scores*) are generally less reliable than either of the parts. Scholastic ability tests are often used to predict achievement test scores, and the difference (or residual score) is computed between the observed score and the predicted score. Such residual scores are, in general, less reliable than the original observed scores. Another composite score that is usually less reliable than are its parts is a ratio score.

Standard 2.2

The procedures that are used to obtain samples of individuals, groups, or observations for the purpose of estimating reliabilities and standard errors of measurement, as well as the nature of the populations involved, should be described. The numbers of individuals in each sample that are used to obtain the estimates, score means, and standard deviations should also be reported. ***(Primary)***

Standard 2.6

Standard 2.7

Standard 2.3

Each method of estimating a reliability that is reported should be defined clearly and expressed in terms of variance components, correlation coefficients, standard errors of measurement, percentages of correct decisions, or equivalent statistics. The conditions under which the reliability estimate was obtained and the situations to which it may be applicable should also be explained clearly. ***(Primary)***

59X 104

Comment:
Because there are many ways of estimating reliability, each influenced by different sources of measurement error, it is unacceptable to say simply, "The reliability of test X is .90." A better statement is, "Based on the correlation between alternate test forms A and C administered on successive days to a sample of 100 tenth-grade students from a middle-class suburban public school in New York, the alternate form reliability is estimated to be .90, with an approximate 95% confidence interval of (.85-.93)."

Standard 2.4

If reliability coefficients are adjusted for restriction of range, both the adjusted and unadjusted coefficients should be reported together with the standard deviations of the group actually tested and of the group for which adjusted estimates are presented. *(Primary)*

Standard 2.5

Estimates of reliability that are based on alternate forms of a test administered to the same sample of individuals on two separate occasions should indicate the order in which forms were administered, the interval between administrations, and a rationale for choosing that interval. Means and standard deviations obtained from both forms should be provided, as well as standard errors of measurement and the estimate of the alternate-form reliability. *(Primary)*

Comment:
An observed score typically represents the performance of a test taker during a particular period of time, which may be a few years, several months, or only an hour or so (as in measures of mood, for example). Evidence should be provided for the consistency of the information obtained by independent measurements on two or more occasions during the period in which test interpreters are likely to regard a person's score as stable.

In some cases it may be advisable to obtain scores on more than two occasions, particularly if considerable instability is expected. Where parallel forms are used in an investigation of stability, it should be recognized that content differences between forms, as well as instability, contribute to the error variance. Estimates of stability based on a retest with the same form, however, may be spuriously inflated due to the effects of memory.

Standard 2.6

Coefficients based on internal analysis should not be interpreted as substitutes for alternate-form reliability or estimates of stability over time unless other evidence supports that interpretation in a particular context. *(Primary)*

Standard 2.7

Procedures known to yield inflated estimates of reliability for speeded tests should not be used to estimate the reliability of a highly speeded test. *(Primary)*

Comment:
For example, split-half coefficients that are obtained from scoring odd and even numbered test items separately yield an inflated estimate for a highly speeded test and are thus inappropriate.

60X
105
BEST COPY AVAILABLE

3. Test Development and Revision

Background

This chapter covers issues of general concern to test developers, emphasizing how test development research can provide the basis for examining issues discussed in other chapters. Test developers have a responsibility to provide evidence regarding reliability and validity for stated testing purposes, as well as manuals and norms, when appropriate, to guide proper interpretation. They also need to anticipate how their tests will be used and misused, to do research that helps distinguish proper from improper uses, and to design tests and accompanying materials in ways that promote proper uses. The mode of presentation, that is, manuals or other materials, is not specified in many of the following standards; however, the test developer, publisher, or sponsor has a responsibility to present information in a readily available form, with summaries and interpretations, to facilitate test review and evaluation.

Although it is concerned with strengthening current testing practices, the *Standards* is also intended to encourage the development of new and improved tests, so that the contributions of tests and testing to society can be extended. Advances in testing stem from research in a variety of areas. For example, some experiments in cognitive psychology are being transformed into faceted diagnostic assessment batteries; physiological and neuropsychological measures are being investigated as potential selection and classification devices; learning sample tests and learning style inventories are being used to prescribe educational treatments; and computerized adaptive and interactive testing, multimedia test presentations, and computerized interpretations are being used increasingly. In the *Standards* an attempt is made to anticipate problems posed by such developments and to facilitate advantages they offer.

The standards in this chapter cover test and item specifications, item analysis and selection procedures, and the evaluation of test designs for intended uses. Some special standards applicable to particular types of tests, including computerized tests, are also included.

Standard 3.1

Tests and testing programs should be developed on a sound scientific basis. Test developers should compile the evidence bearing on a test, decide which information is needed prior to test publication or distribution and which information can be provided later, and conduct any needed research. ***(Primary)***

Standard 3.2

The specifications used in constructing items or selecting observations and in designing the test instrument as a whole should be stated clearly. The definition of a universe or domain that is used for constructing or selecting items should be described. When, for reasons of security, sample copies of a test are unavailable for inspection, the descriptive information should include a representative item identified with each major cell in the classification or domain definition. When item difficulty is a facet of such a system, items representative of the difficulty levels should be provided. ***(Conditional)***

Comment:
Test specifications sometimes indicate that a test is criterion-referenced as opposed to norm-referenced, and this practice has led to some confusion. In norm-referenced interpretations, a score (for an individual or for a definable group) is compared with distributions of scores for other

61K

106

individuals or groups. In criterion-referenced interpretations, the score is taken to reflect directly a level of competence in some defined criterion domain. Although tests built with different reference specifications may differ in various ways, the interpretation of the test scores--not the test itself--is norm-referenced or criterion-referenced. Thus some norm-referenced tests can be interpreted in criterion-referenced ways and vice versa.

The adequacy and usefulness of criterion-referenced interpretations depend on the rigor with which the behavioral domain represented by the test has been defined. Such interpretations are intended to describe the status of individuals or groups with respect to one or more behavioral domains, and it is the domain definition that provides the primary reference for interpretation of scores and for judging the adequacy of the test. The domain definition should be sufficiently detailed and delimiting to show clearly what facets of behavior are included and what facets are excluded in the domain. Within the domain, the classification system adopted should show clearly what and how many facets of behavior the domain comprises.

Standard 3.6

Standard 3.7

Standard 3.8

Standard 3.3

Domain definitions and the test specifications should be sufficiently clear so that knowledgeable experts can judge the relations of items to the domains they represent. ***(Primary)***

Standard 3.4

When test items relate to a course of training or study, a curriculum, a textbook, or packaged instruction, the manual or other reports should include an identification and description of the course or instructional materials and should indicate the year in which these materials were prepared. ***(Secondary)***

Standard 3.5

When selecting the type and content of items for tests and inventories, test developers should consider the content and type in relation to cultural backgrounds and prior experiences of the variety of ethnic, cultural, age, and gender groups represented in the intended population of test takers. ***(Conditional)***

Comment:
For some kinds of test content, cultural background factors are irrelevant, as in simple numerical tests of arithmetic skills or in some employment tests. When the relevance of such factors is in doubt, test developers might establish a review process using expert judges both to select item material and to eliminate material likely to be inappropriate or offensive for groups in the test-taking population. Logical exceptions to this standard are tests of English designed for and used with diverse foreign populations and tests of foreign languages for English-speaking populations.

At various points in test development, empirical procedures may be needed. Such procedures may be needed, for example, when constructing interest inventories, in which differential item response rates may exist for different gender, ethnic, and educational groups. Differential response rates do not necessarily invalidate such items or scales based on them. However, the developer's aim should be to maximize scale validity and, within this constraint, the developer should strive to minimize the potential misrepresentation of interests for major groups in the population that is served.

Standard 3.9

Standard 3.10

Standard 3.11

62X

107

of such strategies by all test takers should be encouraged if their effect facilitates performance and discouraged if their effect interferes with performance. *(Primary)*

Comment:
Test-taking strategies, such as guessing, skipping all doubtful items, or skipping and then returning to doubtful items as time allows, can influence test scores positively or negatively depending on the scoring system used and aspects of item and test design such as speededness or the number of response alternatives provided in multiple-choice items. Differential use of such strategies by test takers can result in reduced test reliability and validity. The goal of test directions, therefore, should be to convey information on the possible effectiveness of various strategies and thus provide all test takers an equal opportunity to perform optimally.

Standard 3.17

Standard 3.18

Standard 3.12

Probable sources of variance that would confound the construct or domain definitions underlying the test should be investigated by the test developer, and the implications of the results for test design, interpretation, and use should be presented in the technical manual or in supplementary reports. In general, evidence from research should be provided to justify the use of novel item or test formats. ***(Secondary)***

Standard 3.19

Standard 3.13

For tests that impose strict time limits, test development research should examine the degree to which scores include a speed component and evaluate the appropriateness of that component, given the constructs or content the test is designed to measure. ***(Conditional)***

Standard 3.20

Standard 3.14

The sensitivity of test performance to improvement with practice, coaching, or brief instruction should be studied as part of developmental research, especially on performance tests that use an unfamiliar response mode, such as computer-administered tests. A test that is designed to measure learning from practice, coaching, or instruction should be shown to do so, and a test that is designed to be unaffected by these forms of learning should be shown to be so. Materials to aid in score interpretation should summarize evidence derived from such studies to indicate the degree to which improvement with practice or coaching can be expected. ***(Secondary)***

Standard 3.21

Standard 3.15

For interest or personality measures intended for selection or placement purposes, evidence should be presented on the extent to which scores are susceptible to an attempt by test takers to present false or unduly favorable pictures of themselves. ***(Secondary)***

Standard 3.22

Standard 3.16

The score report forms and instructional materials for a test, including computerized reports and materials, should facilitate appropriate interpretations. ***(Primary)***

63X

Comment:
This standard is particularly important in the case of computer programs or computerized reports provided to test takers.

Standard 3.17 **If a short form of a test is prepared by reducing the number of items or organizing portions of a test into a separate form, empirical data or a theoretical rationale should be provided to estimate the reliability of each short form and its correlation with the standard form.** ***(Primary)***

Standard 3.18 **A test should be amended or revised when new research data, significant changes in the domain represented, or new conditions of test use and interpretation make the test inappropriate for its intended uses. An apparently old test that remains useful need not be withdrawn or revised simply because of the passage of time. But it is the responsibility of test developers and test publishers to monitor changing conditions and to amend, revise, or withdraw the test as indicated.** ***(Primary)***

Standard 3.19 **Tests should not be titled or advertised as "revised" unless they have been revised in significant ways. A phrase such as "with minor modification" should be used when the test has been modified in minor ways. The score scale should be adjusted to account for these modifications.** ***(Primary)***

Standard 3.20 **If a test or part of a test is intended for research use only and is not distributed for operational use, this fact should be displayed prominently in any materials provided for interpreting individual scores.** ***(Primary)***

Standard 3.21 **The directions for test administration should be presented with sufficient clarity and emphasis so that it is possible to approximate for others the administrative conditions under which the norms and the data on reliability and validity were obtained.** ***(Primary)***

Comment:
Because people administering tests in schools, industry, and in other settings sometimes may not understand the need to follow instructions closely, it is necessary that test administrators receive detailed and insistent instruction on this point.

Standard 3.22 **The directions presented to a test taker should be detailed enough so that test takers can respond to a task in the manner that the test developer intends. When appropriate, sample material and practice or sample questions should be provided.** ***(Primary)***

Comment:
For example, in a personality inventory it may be intended that test takers give the first response that occurs to them. Such an expectation

BEST COPY AVAILABLE

8. Educational Testing and Psychological Testing in the Schools

Background

Testing in education is pervasive. From pre-kindergarten readiness assessments to professional specialty licensing and certification, students participate in a continuing testing and evaluation process designed to monitor their progress, to provide a basis for selection into programs with limited enrollment, and for the award of certificates of qualification. This chapter covers four areas of application: school testing programs, educational certification testing, educational selection, and special education. Chapters 7, 13, and 14 also contain material relevant to educational testing. The use of tests in counseling is covered in Chapter 9 and test use in program evaluation is covered in Chapter 12.

School Testing Programs

Since the early part of this century, school testing programs have been integral to elementary and secondary education. Each year millions of students in thousands of public and private schools take group standardized tests of ability and achievement chosen and administered by their schools, districts, or states. These school testing programs provide local test users with information about the ability and achievement levels of individual students and of groups of students aggregated at various levels.

The test results are used by school administrators, teachers, parents, students, various citizen groups, and the media. The results of carefully selected, appropriate tests, when interpreted properly, can provide administrators with pertinent information about the general academic development and level of functioning of individual students, thereby helping to provide students with appropriate instruction and resources. Test scores can help teachers, students, and parents identify the specific academic strengths of a student on which to build, as well as the specific less-developed areas in need of remediation and special attention.

Other people use test scores as part of formal or informal evaluations of the effectiveness of the school, district, or state in educating students. Often test results on a school-by-school basis are reported by the local news media without comment or explanation of the possible reasons for any differences that may exist between schools within the district. This practice tends to contribute significantly to the misinterpretation and misuse of test results.

The group of users associated with school testing programs is perhaps the largest and most heterogeneous of all user groups. These diverse users should be provided help in interpreting test results. The job may be made easier by using material developed by publishers, but care must be taken to assure that such materials can be understood and are appropriate for local conditions.

Certification Testing in Elementary and Secondary Education

The use of educational certification tests in both elementary and secondary schools has grown rapidly over the past few years at both the state and local levels. "Educational certification test" is a generic term that applies to many different uses of test results and perhaps obscures the considerable diversity among programs. Students' scores on educational certification tests are used either alone or in conjunction with other criteria to make decisions concerning high school graduation or grade-to-grade promotion, to classify students for remediation, to evaluate the effectiveness of schools, to classify or certify school districts, to allocate compensatory funds or other resources to districts, and to evaluate teachers.

An important use of educational certification tests is in the awarding of high school diplomas. In some jurisdictions, if students cannot pass the test, they may receive a certificate of attendance instead of the regular

65X

170

diploma. In others, students who pass the educational certification test receive an endorsed diploma, and those who fail the test, but have met all of the other graduation requirements, are awarded a regular diploma. In some jurisdictions students may earn the high school diploma by passing a test before completing all courses normally required for graduation, and for many years adults have used tests as an alternate route to the high school credential.

Using test scores for classification or certification decisions (i.e., Does the student need remediation? Should the student be promoted or graduated?) presupposes a consensual body of material upon which all students can be reasonably compared. This body of material to which all students have been exposed consists usually of a set of standard textbooks or curricular material, but systematic instruction and a minimum period of study for all students are also implied.

Most educational certification testing programs use a predetermined cut score to distinguish passing from failing scores. The cut score becomes the linchpin in the decision process. Research has shown that there can be large discrepancies between the cut scores produced by the most common methods of setting cut scores. Therefore, the reliability and validity of decisions and inferences based on cut scores from educational certification tests need to be studied carefully.

The agency that mandates an educational certification program, usually the state legislature or the state or local board of education, frequently functions as both the test developer and the test user. Often the agency enters into a contract with a test company to construct a test according to agreed-upon specifications. In such situations the agency is still technically the test developer, even though an external organization actually builds the test. The agency mandating the test and the test development contractor should collaborate to ensure provision of the documentation that permits an evaluation of the degree to which appropriate standards of test development, validity, and reliability have been followed.

Educational Selection

Admissions requirements vary widely among undergraduate colleges and universities, as well as among graduate and professional schools. Competitive admissions, often involving testing, is also becoming increasingly common in elementary and secondary schools. At the undergraduate level, the degree of selectivity ranges from open-door policies that admit any high school graduate (or applicant with equivalent credentials) to highly selective institutions that require considerable evidence of outstanding academic ability and superior past performance. Not only do the requirements vary from institution to institution but also from one specialization to another within an institution. At the graduate level, requirements may differ for applicants to the same department depending on their proposed areas of specialization.

Despite the diversity in how selection testing is used, there are substantial similarities among selection processes. Several types of information are typically required. These may include past academic record (e.g., transcripts, grade-point average, or rank in class), test scores, letters of recommendation, lists of past accomplishments, and statements by the applicant (e.g., goals, personal description, or a writing sample). Descriptive background data such as gender, age, and racial or ethnic group designation are also frequently requested and may be used for affirmative action or other purposes. This list is not intended to be exhaustive or to apply to all institutions, but it does show some of the range.

Criterion-related evidence is the most common approach to validation in admissions contexts, although content-related and construct-related evidence are important. It should be noted that many institutions select students in order to meet other types of objectives besides that of achieving academic excellence, a fact that needs to be taken into account in assessing the validity of the test application.

66X

BEST COPY AVAILABLE

The results of criterion-related validation studies need to be interpreted in the context of previous results at the individual institution and results obtained at other similar institutions. Results for other similar institutions are most relevant in schools or colleges where the number of students available for a criterion-related validity study is small. Results for a single small sample at a particular school or college may be much less informative than results accumulated across many similar institutions.

A logical analysis of the types of questions and the processes necessary to answer those questions is a necessary part of evaluating the validity of a test. The question of whether coaching alters the meaning of the scores provides a useful example. An interpretation that says that a test measures abilities that are developed over the course of many years and that those abilities change slowly as the result of time and effort would be called into question by evidence of significant changes that result from short-term coaching. In general, if the changes are large, the coaching period is relatively short, and the coaching itself deals with essentially nonacademic tasks such as test-taking skills or anxiety, then the validity of the test interpretation is called into question. On the other hand, evidence that coaching results in changes not only in test performance but also in other indicators of academic performance, such as subsequent grades, would help support the validity of the interpretation, especially if the coaching was relatively long term and dealt with the acquisition of academic skills and knowledge.

Special Education

Tests are used in special education to aid in clarifying the types, bases, and extent of an individual test taker's learning difficulties or school adjustment problems. Ultimately, the test results are used in planning an individual education program for the student, sometimes including placement in a special school or classroom. A variety of types of tests is used in special education, including, but not limited to, learning aptitude tests, group or individual achievement tests, tests of specific skills thought to be basic to school learning (e.g., visual-motor integration skills), speech and language tests, vision and hearing tests, personality inventories, behavioral observations, and projective techniques.

In special education, tests are selected, administered, and interpreted by school psychologists, classroom teachers, special educators, and other professionals, such as speech pathologists and physical therapists. This diverse group of test users includes professionals with varying levels of training in measurement and evaluation and with varying degrees of technical expertise in testing. When test users in special education have little or no training in measurement principles, the risk of test misuse is high.

Legislation now requires school officials to evaluate large numbers of children, including children with whom they have not typically had assessment experience--children with low-incidence severe handicaps, preschool children, and people 18 to 21 years of age. This mandated increase in testing, the pressure of time (evaluations must be completed within a specific number of working days after referral), and economic implications (the school is responsible for whatever special education services the evaluation results suggest are needed) have created pressures toward expediency in testing practices. For example, there may be administrative pressure to use less expensive, less time-consuming, or more readily available testing procedures than the test administrator believes are warranted. There may also be pressures not to look for, or not to find, problems that require expensive services, and this may affect, to some extent, the interpretation of test results. In addition, pressures may lead to the use of available but inadequately trained staff to evaluate populations of children with whom they have not previously had experience. Although school staff may be knowledgeable about the assessment of mildly and moderately disabled children of ages 5 to 18, they may not be trained or experienced in the evaluation of the younger, older, or more severely handicapped children who must now be evaluated according to law.

BEST COPY AVAILABLE

Adhering to professional testing standards in special education is increasingly important in the face of these pressures toward expediency. Strict adherence is necessary in those situations in which test-based decisions will have substantial impact on a child's education and life--situations in which special class or school placement is at issue.

Service Delivery to Regular Education

School psychologists employ a wide range of individually administered tests in the process of service delivery to students in regular education classes. These services are for students who may not have special education needs, but have behavioral, emotional, and/or learning problems sufficiently intense to frustrate their educational development, and often the educational development of others. Test results are one source of data in the evaluation of such students.

Standard 8.1

Those responsible for school testing programs should ensure that the individuals who administer the tests are properly instructed in the appropriate test administration procedures and that they understand the importance of adhering to the directions for administration that are provided by the test developer. ***(Primary)***

Standard 8.2

Those responsible for school testing programs should ensure that the individuals who use the test scores within the school context are properly instructed in the appropriate methods for interpreting test scores. ***(Primary)***

Comment:
The interpretation of some test results is sufficiently complex to require that the user have relevant psychological training and experience. Examples of such tests include personality inventories, projective techniques, and neuropsychological tests. Administering and interpreting individually administered intelligence tests also requires extensive training and experience.

Standard 8.3

If test results are used in making statements about the differences between aptitude and achievement for an individual student, any educational decision based on these differences should take into account the overlap between the constructs and the reliability or standard error of the difference score. ***(Primary)***

Comment:
It should not be assumed that, because the words "aptitude" or "ability" are used in the title of a test, it measures a construct distinct from what is measured by an "achievement" test.

Standard 8.4

When a test is to be used to certify the successful completion of a given level of education, either grade-to-grade promotion or high school graduation, both the test domain and the instructional domain at the given level of education should be described in sufficient detail, without compromising test security, so that the agreement between the test domain and the content domain can be evaluated. ***(Primary)***

68X

Standard 8.5 When a test is developed by a state or local district to be used for student promotion, graduation, or classification decisions, user's guides, or technical reports should be developed and disseminated. *(Conditional)*

Comment:
An agency that develops a certification or classification test has the same obligation to supply a manual and technical reports as does a commercial test publisher. A test that is widely used throughout a jurisdiction, even though not published or sold, requires a technical manual so that it can be properly used and evaluated. In smaller testing programs, dissemination may be limited to summary statements, provided that detailed analyses are made available on request.

Standard 8.6 Results from certification tests should be reported promptly to all appropriate parties, including students, parents, and teachers. The report should contain a description of the test, what is measured, the conclusions and decisions that are based on the test results, the obtained score, information on how to interpret the reported score, and any cut score used for classification. *(Primary)*

Standard 8.7 When a test is used to make decisions about student promotion or graduation, there should be evidence that the test covers only the specific or generalized knowledge, skills, and abilities that students have had the opportunity to learn. *(Primary)*

Standard 8.8 Students who must demonstrate mastery of certain skills or knowledge before being promoted or granted a diploma should have multiple opportunities to demonstrate the skills. *(Primary)*

Standard 8.9 Relationships between predictors and criterion measures that are used in educational admissions should be described by regression equations and associated standard errors of estimate or by expectancy tables in addition to correlation coefficients. *(Primary)*

Standard 8.10 The possibility that differential prediction exists in educational selection for selected groups should be investigated where there is prior evidence to suggest that positive results may be found and where sample sizes are adequate. *(Conditional)*

Comment:
The difficulty posed by small samples is particularly acute for questions of differential prediction among some groups. Native Americans, for example, form such a small fraction of the overall population that few schools can be expected to have enough students for an adequate differential prediction study. Thus, the only feasible way of addressing the question is through cooperative efforts by many institutions that allow combining information across institutions.

BEST COPY AVAILABLE

Standard 8.5 When a test is developed by a state or local district to be used for student promotion, graduation, or classification decisions, user's guides, or technical reports should be developed and disseminated. *(Conditional)*

Comment:
An agency that develops a certification or classification test has the same obligation to supply a manual and technical reports as does a commercial test publisher. A test that is widely used throughout a jurisdiction, even though not published or sold, requires a technical manual so that it can be properly used and evaluated. In smaller testing programs, dissemination may be limited to summary statements, provided that detailed analyses are made available on request.

Standard 8.6 Results from certification tests should be reported promptly to all appropriate parties, including students, parents, and teachers. The report should contain a description of the test, what is measured, the conclusions and decisions that are based on the test results, the obtained score, information on how to interpret the reported score, and any cut score used for classification. *(Primary)*

Standard 8.7 When a test is used to make decisions about student promotion or graduation, there should be evidence that the test covers only the specific or generalized knowledge, skills, and abilities that students have had the opportunity to learn. *(Primary)*

Standard 8.8 Students who must demonstrate mastery of certain skills or knowledge before being promoted or granted a diploma should have multiple opportunities to demonstrate the skills. *(Primary)*

Standard 8.9 Relationships between predictors and criterion measures that are used in educational admissions should be described by regression equations and associated standard errors of estimate or by expectancy tables in addition to correlation coefficients. *(Primary)*

Standard 8.10 The possibility that differential prediction exists in educational selection for selected groups should be investigated where there is prior evidence to suggest that positive results may be found and where sample sizes are adequate. *(Conditional)*

Comment:
The difficulty posed by small samples is particularly acute for questions of differential prediction among some groups. Native Americans, for example, form such a small fraction of the overall population that few schools can be expected to have enough students for an adequate differential prediction study. Thus, the only feasible way of addressing the question is through cooperative efforts by many institutions that allow combining information across institutions.

70X

Standard 8.11

Test users should not imply that empirical evidence exists for a relationship among particular test results, prescribed educational plans, and desired student outcomes unless such evidence is available. ***(Primary)***

Comment:
Test results in special education are often used to develop specific educational objectives and instructional strategies that are assumed to remediate a student's educational deficits or to enable the student to compensate for them. This assumes a relationship among test results and instructional technologies that may not have been demonstrated. In some cases there is limited empirical evidence for a relationship among test results, instructional strategies, and student achievement outcomes.

When evidence supporting the utility of testing procedures for instructional purposes is lacking, test users can stress the tentative nature of the recommendations they provide and encourage teachers and others to weigh the usefulness of that information in light of additional available data.

Backgroun

Standard 8.12

In elementary or secondary education, a decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score. Other relevant information for the decision should also be taken into account by the professionals making the decision. ***(Primary)***

Comment:
A student should not be placed in special classes or schools, for example, solely on the basis of an ability test score. Other information about the student's ability to learn, such as observations by teachers or parents, should also play a part in such decisions.

13. Testing Linguistic Minorities

Background

For a non-native English speaker and for a speaker of some dialects of English, every test given in English becomes, in part, a language or literacy test. Therefore, testing individuals who have not had substantial exposure to English as it is used in tests presents special challenges. Test results may not reflect accurately the abilities and competencies being measured if test performance depends on these test takers' knowledge of English. Thus special attention may be needed in many aspects of test development, administration, interpretation, and decision making. English language proficiency tests, if appropriately designed and used, are an obvious exception to this concern because they are intended to measure familiarity with English as is required in educational settings.

Individuals who are familiar with two or more languages can vary considerably in their ability to speak, write, comprehend aurally, and read in each language. These abilities are affected by the social or functional situations of communication. Some people may develop socially and culturally acceptable ways of speaking that intermix two or even three languages simultaneously. Some individuals familiar with two languages may perform more slowly, less efficiently, and at times, less accurately, on problem-solving tasks that are administered in the less familiar language. It is important, therefore, to take language background into account in developing, selecting, and administering tests and in interpreting test performance.

In Chapter 1 of the *Standards*, validity is discussed at length. The present chapter extends this discussion, emphasizing the importance of recognizing the limits of interpretations drawn from tests developed without due consideration for the influence of the linguistic characteristics of some test takers.

Tests in Other Languages

Testing in the language of the test takers may sometimes be appropriate. However, there are a number of hazards to be avoided in dual-language tests. One cannot assume that translation produces a version of the test that is equivalent in content, difficulty level, reliability, and validity. Psychometric properties cannot be assumed to be comparable across languages or dialects. Many words have different frequency rates or difficulty levels in different languages or dialects. Therefore, words in two languages that appear to be close in meaning may differ radically in other ways important for the test use intended. Additionally, test content may be inappropriate in a translated version. For example, a test of reading skills in English that is translated to serve as a test of reading skills in Spanish may include content not equally meaningful to Spanish-speaking students.

Language Proficiency Testing

Language tests that can assist in appropriate educational program placement are needed in order to accommodate the large number of people in U.S. schools who have not had sufficient opportunity to learn the English used in schools. The need is particularly pressing in the education of young children but is important also in adult education. In some situations giving tests both in English and in the native language may be necessary to determine the kind of instruction likely to be most beneficial.

Because students are expected to acquire proficiency in English that is appropriate to their ages and educational levels, tests suitable for assessing their progress are needed. Some tests that are prepared for students of English as a foreign language may not be useful if they place insufficient emphasis on the assessment of important listening and speaking skills. Measures of competency in all relevant English language skills (commun-

72X 177

BEST COPY AVAILABLE

icative competence, literacy, grammar, pronunciation, and comprehension) are likely to be valuable.

Observing students' speech in naturalistic situations can provide additional information about their proficiency in a language. This may not, however, be sufficient to judge their ability to function in that language in formal situations, such as in the classroom. For example, it is not appropriate to base judgments of a child's ability to benefit from instruction in English solely on language fluency observed in playground speech.

In general, there are special difficulties attendant upon the use of a test with individuals who have not had an adequate opportunity to learn the language of the test. A broader than normal range of tests and observations may be desirable if important decisions are to be based on the test results.

Standard

Standard

Individual Testing in Schools

Behavior that may appear eccentric or that may be judged negatively in one culture may be appropriate in another. For example, children from some cultures may be reluctant to speak in elaborate language to adults. Children reared in some cultures may be trained to speak to adults only in response to specific questions or with formulaic utterances. Thus, in a testing situation such children may respond to an adult who is probing for elaborate speech with only short phrases or by shrugging their shoulders.

High levels of verbal output is another example of behavior that may have different values across cultures. One group may judge verbosity or rapid speech as rude, whereas another may regard those speech patterns as indications of high mental ability or friendliness. A child from one culture who is evaluated with mores appropriate to another culture may be considered taciturn, withdrawn, or of low mental ability. Resulting interpretations and prescriptions of treatment may be invalid and potentially harmful to the individual being tested.

Standard

Standard

Standard 13.1

For non-native English speakers or for speakers of some dialects of English, testing should be designed to minimize threats to test reliability and validity that may arise from language differences. ***(Primary)***

Standard

Comment:
Some tests are inappropriate for use with linguistic minority members whose knowledge of the test language is questionable. Careful professional judgment is required to determine when language differences are relevant. Furthermore, the means by which test users meet this standard will vary with different testing situations. Test users can judge what means are most appropriate to their particular use. Some examples of ways in which this standard might be addressed are as follows:

Standard

Standard

1. In some group testing situations where many test takers typically come from a particular linguistic minority, the test administration might profitably be conducted by personnel specially trained to interact with members of that group.
2. In many individual assessment situations, such as in clinical testing, a specially trained test administrator may be able to use the test taker's native language or bilingual speech to elicit test responses more effectively. Bilingual communication may be particularly appropriate in testing individuals from groups known to be commonly bilingual (e.g., Chamorro-English speakers from Guam).

X72B

icative competence, literacy, grammar, pronunciation, and comprehension) are likely to be valuable.

Observing students' speech in naturalistic situations can provide additional information about their proficiency in a language. This may not, however, be sufficient to judge their ability to function in that language in formal situations, such as in the classroom. For example, it is not appropriate to base judgments of a child's ability to benefit from instruction in English solely on language fluency observed in playground speech.

In general, there are special difficulties attendant upon the use of a test with individuals who have not had an adequate opportunity to learn the language of the test. A broader than normal range of tests and observations may be desirable if important decisions are to be based on the test results.

Standard

Standard

Individual Testing in Schools

Behavior that may appear eccentric or that may be judged negatively in one culture may be appropriate in another. For example, children from some cultures may be reluctant to speak in elaborate language to adults. Children reared in some cultures may be trained to speak to adults only in response to specific questions or with formulaic utterances. Thus, in a testing situation such children may respond to an adult who is probing for elaborate speech with only short phrases or by shrugging their shoulders.

High levels of verbal output is another example of behavior that may have different values across cultures. One group may judge verbosity or rapid speech as rude, whereas another may regard those speech patterns as indications of high mental ability or friendliness. A child from one culture who is evaluated with mores appropriate to another culture may be considered taciturn, withdrawn, or of low mental ability. Resulting interpretations and prescriptions of treatment may be invalid and potentially harmful to the individual being tested.

Standard

Standard 13.1

For non-native English speakers or for speakers of some dialects of English, testing should be designed to minimize threats to test reliability and validity that may arise from language differences. ***(Primary)***

Standard 1

Comment:
Some tests are inappropriate for use with linguistic minority members whose knowledge of the test language is questionable. Careful professional judgment is required to determine when language differences are relevant. Furthermore, the means by which test users meet this standard will vary with different testing situations. Test users can judge what means are most appropriate to their particular use. Some examples of ways in which this standard might be addressed are as follows:

Standard 1

1. In some group testing situations where many test takers typically come from a particular linguistic minority, the test administration might profitably be conducted by personnel specially trained to interact with members of that group.
2. In many individual assessment situations, such as in clinical testing, a specially trained test administrator may be able to use the test taker's native language or bilingual speech to elicit test responses more effectively. Bilingual communication may be particularly appropriate in testing individuals from groups known to be commonly bilingual (e.g., Chamorro-English speakers from Guam).

Standard 1

3. In individual assessments, the test administrator may also need to be able to take into account language behavior that is considered socially appropriate in the culture of the test taker. For example, slowness or rapidity of response is influenced by culturally learned speech patterns that are known to vary across linguistic groups.

Standard 13.2 **Linguistic modifications recommmended by test publishers should be described in detail in the test manual.** ***(Primary)***

Standard 13.3 **When a test is recommended for use with linguistically diverse test takers, test developers and publishers should provide the information necessary for appropriate test use and interpretation.** ***(Primary)***

Comment:
Test developers should include in test manuals and in instructions for interpretation explicit statements about the applicability of the test with individuals who are not native speakers of English. However, it should be recognized that test developers and publishers will seldom find it feasible to conduct studies specific to the large number of linguistic groups in this country.

Standard 13.4 **When a test is translated from one language or dialect to another, its reliability and validity for the uses intended in the linguistic groups to be tested should be established.** ***(Primary)***

Comment:
For example, if a test is translated into Spanish for use with Mexican and Puerto Rican populations, its reliability and validity should be established with members of each of these groups.

Standard 13.5 **In employment, licensing, and certification testing, the English language proficiency level of the test should not exceed that appropriate to the relevant occupation or profession.** ***(Primary)***

Standard 13.6 **When it is intended that the two versions of dual-language tests be comparable, evidence of test comparability should be reported.** ***(Primary)***

Standard 13.7 **English language proficiency should not be determined solely with tests that demand only a single linguistic skill.** ***(Primary)***

Comment:
For example, a multiple-choice, pencil-and-paper test on vocabulary does not indicate how well a person understands the language when spoken nor how well the person speaks the language. However, the test score might be helpful in determining how well a person understands some aspects of the written language. In making placement decisions, for example, a more complete range of language skills needs to be assessed.

BEST COPY AVAILABLE

14. Testing People Who Have Handicapping Conditions

Background

Tests are administered to people who have handicapping conditions in a variety of settings and for diverse purposes. There are a number of modifications of tests and test administration procedures that make it possible for people with certain handicapping conditions to take tests developed originally for the general population. Some modified tests, with accompanying research, have been made available by the major national testing programs for a number of years. Although the development of tests and testing procedures for such people is encouraged by the *Standards*, it should be noted that all relevant individual standards given elsewhere in this document are fully applicable to the testing applications considered in this chapter.

Some of the modifications in the way a test is administered alter the medium in which the test instructions and questions are presented to the test takers. For visually impaired people a variety of modifications may be needed. The test booklet may be produced in large print, high-quality regular print, or braille, or the test may be tape-recorded or read aloud to the test taker.

Many hearing-impaired individuals, especially the prelingually deaf, have difficulty in understanding written as well as spoken language; therefore, the intelligibility of the instructions for tests, whether written or spoken, should be considered when tests are modified for hearing-impaired test takers. Modifications of test administration for deaf and hearing-impaired people often include having an interpreter who signs or otherwise interprets the test instructions and, occasionally, the test questions.

The method used to record a response may also need to be modified. Test takers who cannot record their answers to test questions are assisted most commonly by a person who writes or marks the answers. Other ways of obtaining a response include having the respondent use a tape recorder, a typewriter, or a braillewriter. A test may have to be modified to allow a test taker to point to the response of his or her choice.

Nearly all national testing programs that provide modified test procedures for handicapped people provide additional time to take the test. Reading braille and using a cassette recorder or a reader take longer than reading regular print. Reading large type may or may not be more time consuming, depending on the layout of the material and on the nature and severity of the impairment.

Although modifications in the time allowed for tests are considered among the appropriate test options, there are few data available to support any conclusions about the effects of modifications in time, number of sittings, or number of recesses on the test results. Furthermore, little is known about how much time people with various handicapping conditions actually need because records of the time actually used are rare, and empirical studies to set time limits are even more rare.

Changes in test content are sometimes required for test takers with visual or hearing impairments. Items may be unnecessarily difficult for visually impaired people if they use visual stimuli to measure knowledge acquired through other senses. This problem can be identified and corrected by simply reviewing the items, spotting the offenders, and substituting nonvisual stimuli. Because the substitutions may alter other characteristics of the items, however, the modified items should be tried out before they are used in operational testing situations. In certain situations the test may also cause problems if it measures knowledge, skills, or concepts learned primarily through vision.

75X 191

BEST COPY AVAILABLE

Verbal tests may create more severe problems for test takers who are prelingually deaf than for those with visual impairments. However, finding appropriate nonverbal tests to measure the same abilities or to predict the same behavior may be extremely difficult. Although this is a testing problem, it reflects more fundamental difficulties in understanding the nature of abilities, what abilities are needed in certain situations, and what existing abilities may compensate for impaired abilities in certain circumstances.

Standard 11.

Many of the modifications in the ways tests are administered for handicapped people necessitate that the tests be given individually rather than to groups of respondents. The reasons for having an individual administration include the absence of a practical or convenient way to use a group administration, the desire not to interfere with others taking a test in a group, and the desire to reduce the anxiety handicapped people may have about the test. Some additional alterations may be required: for example, changing the location of the standard testing site if it is not accessible to people in wheelchairs; providing tables or chairs that make test takers with certain physical disabilities more comfortable; and altering lighting conditions and associated space needs for people with some visual impairments.

Standard 11.

Standard 11.

Despite the history of attempts to modify tests for handicapped people, significant problems remain. First, there have been few empirical investigations of the effects of special accommodations on the resulting scores or on their reliability and validity. Strictly speaking, unless it has been demonstrated that the psychometric properties of a test, or type of test, are not altered significantly by some modification, the claims made for the test by its author or publisher cannot be generalized to the modified version. The major reason for the lack of research is the relatively small number of handicapped test takers. For example, there are usually not enough students with handicapping conditions entering one school in any given year to conduct the type of validation study that is usually conducted for college admission tests.

Standard 11.

Although modifying tests for individuals with handicapping conditions is generally regarded as desirable, sometimes some very basic, unanswered questions should be confronted. When tests are administered to people with handicapping conditions, particularly those handicaps that affect cognitive functioning, a relevant question is whether the modified test measures the same constructs. Do changes in the medium of expression affect cognitive functioning and the meaning of responses?

Of all the aspects of testing people who have handicapping conditions, reporting test scores has created the most heated debate. Many test developers have argued that reporting scores from nonstandard test administrations without special identification (often called "flagging" of test scores) violates professional principles, misleads test users, and perhaps even harms handicapped test takers whose scores do not accurately reflect their abilities. Handicapped people, on the other hand, have generally said that to identify their scores as resulting from nonstandard administrations and in so doing to identify them as handicapped is to deny them the opportunity to compete on the same grounds as nonhandicapped test takers, that is, to treat them inequitably. Until test scores can be demonstrated to be comparable in some widely accepted sense, there is little hope of happily resolving from all perspectives the issue of reporting scores with or without special identification. Professional and ethical considerations should be weighed to arrive at a solution, either as an interim measure or as continuing policy.

Standard 11.

test takers who are
ts. However, finding
ies or to predict the
h this is a testing
derstanding the na-
ituations, and what
s in certain circum-

ninistered for hand-
idually rather than
lividual administra-
way to use a group
s taking a test in a
ed people may have
quired: for example,
is not accessible to
ake test takers with
tering lighting con-
visual impairments.

handicapped people,
ew empirical inves-
the resulting scores
unless it has been
t, or type of test, are
ns made for the test
ie modified version.
ely small number of
lly not enough stu-
l in any given year
onducted for college

dicapping conditions
y basic, unanswered
stered to people with
that affect cognitive
d test measures the
sion affect cognitive

licapping conditions,
ite. Many test devel-
lard test administra-
ging" of test scores)
, and perhaps even
urately reflect their
generally said that
administrations and
hem the opportunity
st takers, that is, to
onstrated to be com-
hope of happily re-
ores with or without
derations should be
measure or as con-

Standard 14.1

People who modify tests for handicapped people should have available to them psychometric expertise for so doing. In addition, they should have available to them knowledge of the effects of various handicapping conditions on test performance, acquired either from their own training or experience or from close consultation with handicapped individuals or those thoroughly familiar with such individuals. ***(Primary)***

Standard 14.2

Until tests have been validated for people who have specific handicapping conditions, test publishers should issue cautionary statements in manuals and elsewhere regarding confidence in interpretations based on such test scores. ***(Primary)***

Standard 14.3

Forms of tests that are modified for people who have various handicapping conditions should generally be pilot tested on people who are similarly handicapped to check the appropriateness and feasibility of the modifications. ***(Conditional)***

Comment:
Although useful guides to modifying tests are available, they do not provide a universal substitute for trying out a modified test or validating the modified version of a test. Even when such tryouts are conducted on samples inadequate to produce norm or validity data, they should be conducted to check the mechanics of the modifications.

Standard 14.4

Interpretive information that accompanies modified tests should include a careful statement of the steps taken to modify tests in order to alert users to changes that are likely to alter the validity of the measure. ***(Conditional)***

Comment:
If empirical evidence of the nature and effects of changes resulting from modifying standard tests is lacking, it is impossible to enumerate significant modifications that are to be documented in manuals. Therefore, test developers should take care to document all changes made and be alert to indications of possible effects of those modifications. Documentation of the procedure used to modify tests will not only aid in the administration and interpretation of the given test but will also inform others who are modifying tests for people with specific handicapping conditions.

Standard 14.5

Empirical procedures should be used whenever possible to establish time limits for modified forms of timed tests rather than simply allowing handicapped test takers a multiple of the standard time. Fatigue should be investigated as a potentially important factor when time limits are extended. ***(Secondary)***

193
77X
BEST COPY AVAILABLE

Standard 14.6

When feasible, the validity and reliability of tests administered to people with various handicapping conditions should be investigated and reported by the agency or publisher that makes the modification. Such investigations should examine the effects of modifications made for people with various handicapping conditions on resulting scores, as well as the effects of administering standard unmodified tests to them. ***(Secondary)***

Comment:
In addition to modifying tests and test administration procedures for people who have handicapping conditions, validating these tests is urgently needed. Validation is the only way to amass knowledge about the usefulness of tests for people with handicapping conditions. The costs of validating these tests should be weighed against those of not having usable information regarding the meanings of scores for handicapped people.

Standard 14.7

Those who use tests and those who interact professionally with potential test takers with handicapping conditions (e.g., high school guidance counselors) should (a) possess the information necessary to make an appropriate selection of alternate measures, (b) have current information regarding the availability of modified forms of the test in question, (c) inform individuals with handicapping conditions, when appropriate, about the existence of modified forms, and (d) make these forms available to test takers when appropriate and feasible. ***(Primary)***

Standard 14.8

In assessing characteristics of individuals with handicapping conditions, the test user should use either regular or special norms for calculating derived scores, depending on the purpose of the testing. Regular norms for the characteristic in question are appropriate when the purpose involves the test taker's functioning relative to the general population. If available, however, special norms should be selected when the test takers' functioning relative to their handicapped peers is at issue. ***(Primary)***

reports with appropriate cautions regarding the possible effects of such modifications on validity. *(Primary)*

Standard 15.5 — Test scoring services should document the procedures that were followed in order to assure accuracy of scoring. The frequency of error should be monitored and reported on request. *(Conditional)*

Standard 15.6 — When the score report may be the basis on which decisions would be made in the near future and a material error is found in test scores or other important information released by a testing organization or other institution, a corrected score report should be distributed as soon as it is practicable. *(Primary)*

Standard 15.7 — Test users should protect the security of test materials. *(Primary)*

Comment:
Those who have test materials under their control should take all steps necessary to assure that only individuals with a legitimate need for access to test materials are able to obtain such access.

Standard 15.8 — In educational admissions and licensing or certification applications, in which important decisions depend on performance on a given test, a means of checking the accuracy of the scoring should be available to test takers. When the test itself and the scoring key cannot be released, some other means of verification should be provided. *(Conditional)*

Standard 15.9 — When test data about a person are retained, both the test protocol and any written report should also be preserved. *(Primary)*

Standard 15.10 — Those responsible for testing programs should provide appropriate interpretations when test score information is released to students, parents, legal representatives, teachers, or the media. The interpretations should describe in simple language what the test covers, what scores mean, common misinterpretations of test scores, and how scores will be used. *(Primary)*

Comment:
Test users should consult the interpretive material prepared by the test developer or publisher and should revise or supplement the material as necessary to present the local and individual results accurately and clearly.

Standard 15.11 — Organizations that maintain test scores on individuals in data files or in an individual's records should develop a clear set of policy guidelines on the duration of retention in an individual's records, availability, and use over time of such scores. *(Primary)*

Comment:
In some instances test scores become obsolete over time and should not be used or be available. In other cases test scores obtained in past years can be extremely useful, for example, in longitudinal assessment. The key issue is the valid use of the information.

Backgroun

Standard 1

Standard 1

Standard 1

79X

16. Protecting the Rights of Test Takers

Background

Certain broad principles regarding access to test scores are now widely accepted. Some technical requirements necessary to satisfy these principles are stated as specific standards in this chapter. The issues of test security and the cancellation of test takers' scores because of testing irregularities are also addressed.

Standard 16.1

Informed consent should be obtained from test takers or their legal representatives before testing is done except (a) when testing without consent is mandated by law or governmental regulation (e.g., statewide testing programs); (b) when testing is conducted as a regular part of school activities (e.g., schoolwide testing programs and participation by schools in norming and research studies); or (c) when consent is clearly implied (e.g., application for employment or educational admissions). When consent is not required, test takers should be informed concerning the testing process. ***(Primary)***

Comment:
Informed consent implies that the test takers or representatives are made aware, in language that they can understand, of the reasons for testing, the type of tests to be used, the intended use and the range of material consequences of the intended use, and what testing information will be released and to whom. When law mandates testing but does not require informed consent, test users should exercise discretion in obtaining informed consent, but test takers should always be given relevant information about a test when it is in their interest to be informed.

Young test takers should receive an explanation of the reasons for testing. Even a child as young as two or three and many mentally retarded test takers can understand a simple explanation as to why they are being tested. For example, an explanation such as "I'm going to ask you to try to do some things so that I can see what you know how to do and what things you could use some more help with" would be understandable to such test takers.

Standard 16.2

In school, clinical, and counseling applications, test users should provide test takers or their legal representative with an appropriate explanation of test results and recommendations made on the basis of test results in a form that they can understand. ***(Primary)***

Comment:
This standard requires both the use of the appropriate language with non-English speaking test takers and the use of conceptually understandable explanations with all types of test takers. Even children and many mentally retarded test takers can understand a simple explanation of test results.

Standard 16.3

Test results identified by the names of individual test takers should not be released to any person or institution without the informed consent of the test taker or an authorized representative unless otherwise required by law. Scores of individuals identified by name should be made

BEST COPY AVAILABLE

80X

available only to those with a legitimate, professional interest in particular cases. *(Primary)*

Standard

Comment:
Information may be provided to researchers if a test taker's anonymity is maintained and the intended use is not inconsistent with the conditions of the test taker's informed consent.

Standard 16.4

In educational, clinical, and counseling applications, when test scores are used to make decisions about individuals, the affected person or legal representative should be able to obtain transmittal of this test score and its interpretation for any appropriate use. *(Secondary)*

Standard 16.5

Test data maintained in data files should be adequately protected from improper disclosure. Use of time-sharing networks, data banks, and other electronic data processing systems should be restricted to situations in which confidentiality can be reasonably assured. *(Primary)*

Standard 16.6

When score reporting includes assigning individuals to categories, the categories chosen should be based on carefully selected criteria. The least stigmatizing labels, consistent with accurate reporting, should always be assigned. *(Primary)*

Standard 16.7

Under certain conditions it may be desirable to cancel a test taker's score or to withhold it because of possible testing irregularities, including suspected misconduct. The type of evidence and procedures to be used to determine that a score should be canceled or withheld should be explained fully to all test takers whose scores are being withheld or canceled. *(Primary)*

Standard 16.8

In educational admissions and licensing and certification applications, when a score report will be delayed beyond a brief investigative period because of possible irregularities such as suspected misconduct, the test taker should be notified, the reason given, and reasonable efforts made to expedite review and to protect the interests of the test taker. *(Primary)*

Standard 16.9

In educational admissions and licensing and certification applications, before a score is canceled or its report is withheld beyond a brief investigative period, test takers should be given advance warning and an opportunity to provide evidence that the score should not be canceled or withheld. All evidence considered in deciding upon the intended action, including evidence that might lead to a contrary decision, should be made available to the test taker on request. *(Primary)*

Comment:
Some testing organizations offer the option of a prompt and free retest or arbitration of disputes.

81X

THE CAPTIVE PSYCHOLOGIST AND THE CAPTIVE PATIENT

Louise Mead Riscalla

The psychologist and members of other professions working with court-referred cases may find themselves in a captive, adversary position, an outgrowth of the American adversary-type judicial system. It is posited that the dilemmas involved in a captive, adversary position represent the personal attitudes of the professional working within the judicial system. Some of the captive reactions of both psychol and patient will be explored witi recognition of the possibility that ers of other professions may respo.. .: a similar manner.

The work of a psychologist as a service to the courts or in correctional settings is varied. This article will focus on psychological assessment which is generally the main reason for referral to psychologists by the courts. The psychologist is required, either individually, or as part of a team, for example, to make determinations of an individual's competency to stand trial, to decide on the basis of evaluation(s) whether a juvenile should remain with his family or be placed in a correctional or other facility, whether or not an offender should be incarcerated or receive probation, and other determinations which affect the daily lives and future of the clients.

The psychologist is in a captive position when he is placed in an adversary role by being administratively compelled to prepare evaluations and handle information in accord with procedures which may not be in the best interest of his professional concern or that of the patient. The psychologist is concerned with the offender's rights as a human being, yet, on the other hand, he is employed by the court or a court facility and is also concerned with institutional policy as well as his responsibility to protect the rights of the community.

The issue of confidentiality has been extensively discussed by Shah (1969, 1970a, 1970b) and others. However, for the psychologist working with the court and other governmental and public facilities, there is often no assured privacy or confidentiality. For example, in New Jersey, the psychologist may be subpoenaed by either side to testify in court. There has been concern with the patient's right to review what is said about him in the evaluation. According to a New Jersey Supreme Court decision, the State of New Jersey v. Horne, Andrews, Blanford, Coleman & Barnes, the patient or his representative has the right to subpoena the report of the psychiatric evaluation, including the report of the psychologist, and to contest the findings. While this decision may be beneficial to the patient, it could also place the psychologist in an adversary position. He can either be used by the court to support a particular position or by the patient against the court.

The APA (1970) position statement, "Psychological Assessment and Public Policy," appeared to be mainly concerned with the rights of the psychologist and institutions where psychologists perform assessments. There was nothing implicit or explicit concerning possible shortcomings on the part of psychologists themselves. Goldman (1970) in his comment on the APA position statement said,

> we know that the applicant does not always see the test or the tester as his friend. And we also know that the real psychologist is not always the ideal, that

> ... psychologists usually use poor or obsolete tests, make mistakes in interpretation, or fail to develop and use adequate norms and validity data [p. 874].

It has apparently been assumed by APA that there is no need to question the psychologist, and that the public is protected because the psychologist who has professional principles and is bound by professional ethics is in control. However, especially with court-referred patients, the "principled psychologist" could very well work to the detriment of himself, his profession, and the public.

With all his refined techniques and seemingly objective approach, the psychologist is still a human being subject to personal reactions to his patients as individuals. Very often the moral standards of the psychologist may be at odds with the patient's offenses, particularly in the examination of sex offenders, homicides, or atrocious assault cases. Regardless of how "objective" a psychologist may try to be or appear, his own attitudes often manifest themselves in some manner. The response of the psychologist to his captive, adversary position is an often overlooked consideration. It is important for the psychologist to be aware of and understand his own attitudes toward his captive, adversary role.

It is possible that psychologists could have anti-Establishment attitudes and unresolved conflicts with authority. Such a psychologist might take sides with the patient with the rationalization of protecting the rights of his patient by joining forces with the patient to fight the Establishment. Consequently, the report could be worded in such a way as to be an attack on the court. The offense would be rationalized on the basis of personality dynamics, a reaction to perceived unsatisfactory environmental conditions such as poverty and racial discrimination, which the psychologist believes require changing, or as a perceived self-defense. There is a danger that the psychologist might remove the individual from responsibility connected with the offense by "blaming" the patient's difficulty on society, which amounts to perceiving the individual as a robot. Such a psychologist might try to instigate and incite other staff members to campaign against the court and society. In short, there is also a danger that the psychologist could use his position to wage his own personal anti-Establishment battle. He takes his captive position personally and reacts to perceived opposition by opposing the system. As a consequence, a power struggle takes place obscuring and obstructing the possibility of constructive, beneficial solutions.

A psychologist who is impressed with his influence over the lives of those referred by the court might assume a punitive, self-righteous, superior attitude toward the patients. His aim is to protect the rights of society at any cost. His reports and recommendations could be so worded that the court would take punitive action. He is likely to make moral judgments on the patients and could reasonably support his judgment documented by test findings. He is often unaware that his interpretations and judgment are governed primarily by his own personal feelings about the patients and the use of assessment evidence to support his own position.

There are psychologists who do not face the implications and are unaware of the circumstances of the adversary position, who consequently do not wish to get involved. Their main concern is with doing their job and collecting a

83X

199

BEST COPY AVAILABLE

salary. Often such a psychologist is afraid of authority. He might have the idea that, if he is forthright and open, he could lose his job. Such a psychologist would be likely to write a neutral, benign report which is meaningless, says nothing, and could be aggreable to either side.

Some psychologists recognize their captive, adversary circumstances, but do not place themselves in such a position. They do not take sides because they are aware of the shortcomings of their instruments and the system that they represent, yet work within it in a positive, constructive manner. If, for example, a psychologist detects an error and finds that he might be in a position in the course of his work or daily living experience to point it out and have the opportunity to offer constructive change, he would do so, but in a way where he does not deliberately offend or trespass on the system or rights of others. He exercises wisdom, caution, is perceptive and alert to the difficulties on all sides. He is concerned with illuminating issues and pointing ways to constructive alternatives which might lead to solutions rather than to taking sides. Consequently, the report is generally clear, relevant, written without jargon as much as possible, and useful to all concerned.

The court-referred individual often does not consider himself in need of help and believes that he is "normal." The possibility of the presence of pathology could be an added social stigma having adverse consequences on his relationship to himself, others, job, and education. Patients are frequently referred, among other reasons, for pretrial assessments. Suspicion, hostility, and fear may be expected from individuals referred for court evaluations.

There is a danger that psychologists might confuse or mistake manifestations of the nonvoluntary patient's captivity with his general makeup. Captive, nonvoluntary patients may, for example, manifest depression, low frustration tolerance, hostility, resentment, resistance, fear, evasiveness, suspicion, inertia, all of which might in part represent captivity reactions. It may be observed that patients have been sent to correctional institutions largely based on "captive" findings, coupled with suspicion and hostility toward the court. It is important for the psychologist to try to understand and contextualize the patient's captive reactions in a meaningful way. It is also important in the evaluation to take these captive manifestations into consideration before rendering a diagnostic impression and recommendations.

It is useful and helpful for the court to know and be aware of the patient's reactions to behavioral and physical control, particularly if the psychologist could offer constructive suggestions to the court regarding the handling of the patient. Fiske (1967) studied the reactions of the testee to tests under standardized conditions and found that individuals show a wide variety of reactions to being tested.

It has been demonstrated through a survey of psychologists working for the courts and court facilities that test results were influenced to some degree by the status of the psychologist's office as part of the court machinery (Naar, 1961). Regardless of how much the psychologist may try to alleviate suspicion, hostility, and evasiveness, the fact that he does represent the court is a factor to be faced openly by both the psychologist and individual referred by the court. In view of the uniqueness of each psychologist and patient, it is impossible to make specific suggestions. However, in general, an atmosphere of openness and honesty

84X

100

should prevail, and it is the psychologist's responsibility to become aware of the degree of his own openness. The circumstances are such that, in a sense, psychologist, patient, and court are all captive. It might be wise to point this out to the patient for purposes of mutual understanding. The patient is generally already aware that certain information must be shared with the court prior to assessment. It is difficult, if not impossible to alleviate the patient's anxiety regarding the psychologist's report to the court, regardless of the nature and amount of information submitted to the court.

Very often test interpretations are made objectively by normative data, diagnostic tables, inferrence, and at times speculation, which, on examination, have little or no relevance to patient as a human being and no ability to distinguish between John Doe as a unique individual and other John Does who commit the same offense. It is helpful to include the patient in test interpretations in order to arrive at a clearer understanding.

Fischer's (1970) work included the patient as a coevaluator with the psychologist in the assessment procedure. The results of the evaluation are shared with the patient before the final report is written; it is a mutual endeavor. This approach has been tried with a number of adult and juvenile offenders and has been found to contribute toward the accuracy of assessments, more realistic recommendations to the court, and has also been of therapeutic value because often patients have gained insight into their behavior. Some therapeutic assistance has been rendered during this approach to assessment (Riscalla, 1970), thus linking therapy with the assessment procedure. It can be observed that if the patient refuses assessment, his refusal is often noted in the psychological report in connection with his behavior.

As a consequence, there is a danger that the court might take punitive action by considering the patient, for example, as "antisocial," based on the court's interpretation of the psychologist's report. There are situations in which the patient does not wish to become involved in assessment, or through pathology or resistance is unable to articulate. It is suggested that the patient be invited to participate, and, if he refuses, his refusal should be taken into consideration in the psychological report and conveyed to the court.

Psychological testing has come into a considerable amount of controversy (i.e., Anastasi, 1967; Brim, 1965; Gross, 1962; Holt, 1967; Messick, 1965; Williams, 1970). The human element involved in testing is difficult, if not impossible, to quantify for validity and reliability purposes. Under the circumstances, it would be advisable that, as part of the procedure, the patient, his representative, and other interested authorized individuals designated by the court and patient have the right to discuss and examine, if necessary, psychological material with the psychologist prior to the submission of the psychological report to legal authorities.

Very often psychologists hide behind the mask of confidentiality of their reports by protecting the best interests of the patients, when actually it is the best interests of the psychologists that are being protected. If for any reason it is advisable to withhold information, the reasons for withholding should be indicated and documented by the psychologist.

The sharing of information is not a "policing procedure," but rather, a means of helping the psychologist to assist

85X

101

others better. It also serves to keep the psychologist alert by writing reports which are understandable to those not familiar with psychological jargon. Sharing of material in assessment and reports could be relevant and meaningful, rather than having it filed away as part of a required, completed legal procedure.

The psychologist often has no way of determining how his findings are used, and follow-up is generally lacking. Ideally it is suggested that the court inform the psychologist of disposition, or have the psychologist or diagnostic team present at court hearings. If there is a difference of opinion, the psychologist should be informed of this difference so that corrective measures could be taken.

It is recognized that the realities of budgetary considerations, heavy caseloads, and shortages of personnel frequently render the "should be's" difficult and at times impossible. What is possible and within the scope of immediate reality, is an ongoing dialogue, so that within the present system there can be understanding and an appreciation of the position of fellow captives. The psychologist could take the initiative as a human being with qualities of warmth, wisdom, and understanding, and use his professional skills to provide realistic alternatives which could conceivably lead to constructive changes. The psychologist would then no longer need to be or see himself as a captive who must take the position as an adversary. The psychologist would be a compatible companion in captive circumstances.

REFERENCES

American Psychological Association. Psychological assessment and public policy. *American Psychologist*, 1970, 25, 264-266.

Anastasi, A. Psychology, psychologists, and psychological testing. *American Psychologist*, 1967, 22, 297-306.

Brim, C., Jr. American attitudes toward intelligence tests. *American Psychologist*, 1965, 20, 125-130.

Fischer, C. The testee as co-evaluator. *Journal of Consulting Psychology*, 1970, 17, 70-76.

Fischer, C. Contextual approach to assessment. *Community Mental Health Journal*, in press.

Fiske, D. The subject reacts to tests. *American Psychologist*, 1967, 22, 287-296.

Goldman, L. Position statement on psychological assessment and public policy. *American Psychologist*, 1970, 25, 874.

Gross, M. *The brain watchers*. New York: American Library, 1962.

Holt, R. Diagnostic testing: Present situation and future prospects. *Journal of Nervous and Mental Disease*, 1967, 144, 444-465.

Messick, S. Personality measurement and the ethics of assessment. *American Psychologist*, 1965, 20, 136-142.

Naar, R. Testing in juvenile courts: A survey. *Journal of Clinical Psychology*, 1961, 17, 210.

Riscalla, L. Crisis therapy with adolescents. *Internal Journal of Offender Therapy*, 1970, 14, 40-43.

Shah, S. Privileged communications, confidentiality, and privacy: Privileged communications. *Professional Psychology*, 1969, 1, 56-69.

Shah, S. Privileged communications, confidentiality, and privacy: Confidentiality. *Professional Psychology*, 1970, 1, 159-164. (a)

Shah, S. Privileged communications, confidentiality, and privacy: Privacy. *Professional Psychology*, 1970, 1, 243-252. (b)

State of New Jersey v. Thomas Horne, A-37; Mailyor Andrews, A-145; Booker T. Blanford, A-146; James Coleman, A-147; Norris Barnes, A-149. Supreme Court of New Jersey, September, 1969 Term.

Williams, R. Danger: Testing and dehumanizing black children. *Clinical Child Psychology Newsletter*, 1970, 9, 5-6.

LOUISE MEAD RISCALLA did her undergraduate work at Upsala College, received an MA from St. Louis University, and is currently enrolled in a PhD program in psychology at the New School for Social Research. Ms. Riscalla is licensed as a psychologist in New Jersey and is involved in parttime private practice of therapy, assessment, and consulting.

86X

102

From: Journal of Clinical Child Psychology, 1974, Vol 3, pp 43-45

Louise Mead Riscalla is a Clinical Psychologist at the New Jersey State Diagnostic Center where her work includes the diagnostic assessment of children and adults. She received an M.A. in psychology and sociology (research) from St. Louis University in 1956, studied rehabilitation counselling at Columbia University, and did further graduate work in psychology at Yeshiva University and the New School for Social Research. Her articles on psychological assessment, attitudes of professionals toward "patients," and rehabilitation have appeared in numerous journals.

MENTAL RETARDATION: FACT OR CONJECTURE?

LOUISE MEAD RISCALLA

Studies reported in the literature indicating the deleterious consequences of diagnostic labeling (Combs and Harper, 1967; Dunn, 1968; Jones, 1972; Meyerowitz, 1962; Sarason and Doris, 1969) and perception of mental retardation as a symptom rather than a syndrome (Carter, 1970; Poser, 1969) provide evidence for questioning whether mental retardation as a concept should continue to be used. Although it has been proposed that the term "mental retardation" be used to describe the problems of an individual rather than the individual himself (Jaslow and Smith, 1972), the term itself as observed in the literature and practice, carries a personal, social, and educational stigma.

Traditionally the term "mental retardation" has been used to designate that an individual functions at a level below that which is expected for his age (Poser, 1969). According to the *Diagnostic and statistical manual of mental disorders*, "Mental retardation refers to subnormal general intellectual functioning which originates during the developmental period and is associated with impairment of either learning and social adjustment or maturation, or both (American Psychiatric Association, 1968, p. 14)." While recognizing that the intelligence quotient should not be the only criterion in diagnosing mental retardation, the manual classifies mental retardation according to intelligence quotients. Thus, there is practically a complete reliance on intelligence quotient scores by school systems, state divisions of mental retardation, courts, child guidance centers, and other agencies or facilities for planning and/or treatment purposes.

The traditional definitions of mental retardation are based on a comparison of an individual with a statistically derived norm which assumes that all individuals develop at the same rate. As a consequence, the uniqueness of each individual is not, in actuality, taken into consideration. Since individuals vary in terms of growth rate, personal attributes, character, etc., it is conceivable that the need for and stress on normative data could create and perpetuate mental retardation. Concepts of mental retardation are based on assumptions about an individual rather than on understanding an individual in the uniqueness of his being and context of his lived world. Limitations are automatically placed on the individual because of the use of concepts, with the consequence that the individual is often limited in his growth by professionals and other people who respond in terms of these preconceived limitations.

The current diagnostic procedures usually consist of a work-up by physicians, social workers, psychologists, educators, etc. with the goal of looking at the complete individual. In spite of the fact that most reputable psychologists do not rely exclusively on measured intelligence quotients but consider factors such as social maturation, the individual still continues to be evaluated primarily on what he is expected to learn academically rather than the kinds of learning involved in daily living. Consequently, what may appear to be mental retardation may, in the final analysis, be a conjecture based on knowledge derived from tests or some other form of norm comparison rather than a reality based on experiences necessary for daily living. Dunn (1968) indicated that current diagnostic procedures..."have probably been doing more harm than good in that they have resulted in disability labels and in that they have grouped children homogeneously in school on the basis of these labels [p. 8]." In the field of medicine, Sterescu (1973) stressed the risks of inaccurate diagnoses and cautioned against making diagnoses too rapidly without considering the consequences of diagnostic labels on the individual.

Diagnosis is geared to finding something wrong with an individual, and it is assumed that once this "wrong" is discerned, then the condition can be corrected. In the final analysis, diagnosis can be considered a form of fault-finding with an individual, when in reality, the error could lie in the diagnosis itself. Dunn (1968) pointed out that "diagnosis tends to stop when something has been found wrong with the child, when the why has either been found or conjectured, and when some justification has been found for recommending placement in a special education class [p. 8]." While Carter (1970) indicated that mental retardation is "not a syndrome itself, but is simply a symptom of some other disease or mental process [p. ix]," current treatment procedures are still focused on symptoms. The term "mental retardation" regardless of how it is used is a destructive, self-fulfilling prophecy because the individual acts in accordance with the way he is treated.

Emotional, physical, and cultural factors interfere with and can lower measured intelligence. Thus, what appears as a symptom of mental retardation may actually be variations of deprivation, and this deprivation is what should be treated. What is needed then is a different perspective, which can be best illustrated by an example.

> Jack was suffering from Mongoloidism which is known to involve severe mental retardation when viewed from the traditional perspective and usually involves long term institutionalization in severe cases, or custodial care at home. Typically, Jack's mother had made the rounds of clinics and resources in quest of help for her son. Jack had no formal education because he was classified as "trainable," but not "educable." He had been placed in a sheltered workshop for awhile where he learned to handle simple tools, and did crafts, woodwork and activities of daily living involving simple, routine tasks. When I first saw Jack, he gave the impression of being a "simple person," accustomed to routine and not tolerant of any change in this routine.
>
> In a discussion with Jack's mother, we shared the view that "retarded children" were essentially no different than others, with the exception that they were deprived in some way. I emphasized that the label "mental retardation" often carries expectations of limitations in the form of "can'ts," such as an inability to learn to read, do arithmetic, household chores, go shopping, etc., thus depriving an individual of the opportunity to develop to his full potential. Even when individuals show interest and ability, there is a tendency not to encourage them to expand because of the danger of building false hopes with resulting frustration. Consequently, with the best of inten-

87X

BEST COPY AVAILABLE

tions, "mental retardation" is perpetuated by inadvertently placing limitations on individuals who often have no choice but to act in accordance with these externally imposed limitations. The "normal" label carries with it expectations that an individual compares to the statistical average.

Since Jack's mother was most frequently with him and had the opportunity to learn more about him than anyone else, I suggested that she try to get to know him as a person without preconceived notions or expectations. Although Jack's father was busy with his job, a similar suggestion was made to him. In practical application, for example, I advised that in the course of the household operation, they count money in front of Jack without anything "put on," and that most likely Jack, displaying childhood curiosity, would spontaneously show interest. They should invite him to participate and teach him about money values by taking their cues from Jack's questioning, observing, and handling various coins, etc. Teaching and participation were tied into the household routine. For the first time, he was given an allowance to spend as he wished, similar to boys of his age. He learned by experience. He also took the garbage out, helped the rest of the family clear the table at meals, and ran some errands in the neighborhood for his parents. He was generally treated as an ordinary member of the family. In matters of discipline, no allowances were made because of Jack's "mental retardation." I encouraged Jack's father to have Jack watch him while he worked and help him with his projects. Combining what he learned at the sheltered workshop with applications at home, Jack was able to help his father do household repairs and make useful items for the house. Some years later, when Jack's father broke an ankle and was incapacitated, Jack was able to do many home repairs independently, thus being an asset rather than a burden to his family. At my last contact, I learned that Jack works as a delivery boy for a druggist, is quite self-sufficient and is a flexible human being. While I have never tested Jack, I think that the issue was not how much measured intelligence he has or what was wrong through repeated physical, psychological, educational, and vocational evaluations, but in seeing Jack as a human being revealed through his daily living experiences. Because we could see what Jack could do in practical situations, living itself provided a more natural, realistic assessment.

There is a tendency to seek specific techniques of working with "retarded" individuals. However, techniques alone are superficial and inauthentic (Riscalla, 1973a). Every individual and experience in life is unique, and a technique which leads to positive results for one client may not for another because the technique is the outcome of the human qualities and perspective of both the professional and client. The perspective used with Jack which has, in principle, been applied to many other individuals and situations was that Jack was perceived without limitations or expectations. The methods used to help Jack were applied on a moment-to-moment, situation-to-situation basis, consistent with human nature and circumstances in life. All personal skills and training were used as frames of reference to assist Jack and his parents. In short, it was a human being-to-human being rather than a professional-to-client form of communication, and a positive approach was used throughout. For example, when an obstacle was encountered, rather than avoid it or impose limitations on Jack by telling him how to handle it or give up in failure, I consulted Jack and asked what he would suggest. If his idea did not work, Jack, his parents and I would experiment with various ways of doing things, taking our cues from Jack, until the obstacle was overcome. Everyone benefited and grew, especially Jack, who overcame his "label."

Psychologists are human beings subject to personal reactions, biases, likes and dislikes. Their errors in judgment have a serious impact on the client which can be reflected in assessments, diagnoses, etc. (Riscalla, 1972, 1973b). Consequently, many clients who are labeled as "retarded" should not be so classified. The human element in the labeling of "retarded" and in psychological testing has been demonstrated in a number of studies (e.g. Mercer, 1972; Thomas, Hertzig, Dryman, and Fernandez, 1971; Watson, 1972). *The standards for development and use of educational and psychological tests* (American Psychological Association, 1973) also recognize the human element in psychological testing and, among other things, recommend the avoidance of descriptive labels. The courts have taken action against mislabeling (examples cited in Vergason, 1973). However, legal action often takes place after the fact, and there is a risk that instead of acting as a deterrent, professionals might try to find loopholes to justify their position in order to continue the same procedures.

It is well known that most individuals do not function at their optimum. Everyone—not only the so-called "retarded"—has inherent intellectual capacities which are obscured by physical, cultural, and emotional factors. Furthermore, theoretical grounds exist for giving up the concept of mental retardation in favor of a perspective which perceives the individual as having a particular mode-of-being (Hora, 1962; Riscalla, 1971), or which perceives the individual in the context of his lived experience (Fischer, 1969, 1970, 1973).

Instead of the concept of mental retardation, an application of these alternative perspectives could perceive an individual, in his lived world, as deprived in some form unique to the particular individual involved. These perspectives would be manifested, for example, by assessing what the individual can do as revealed through living experiences. The provision of tasks of a simple, routine nature is aimed at giving the individual an opportunity for success in achievement. In actuality, such tasks often lead to boredom and impose further limitations. In cutting the individual off from the mainstream of life itself, he is deprived of the opportunity to grow to his potential.

If psychologists respond to a client without preconceived notions derived from labels and concepts and concentrate on enhancing his unqiue developmental pattern, the concept and the term "mental retardation" could be classed as conjecture.

New Jersey State Diagnostic Center
Menlo Park
Edison, New Jersey 08837

REFERENCES

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders.* (2nd ed.) Washington, D.C.: American Psychiatric Association, 1968.

American Psychological Association. *Standards for development and use of educational and psychological tests* (third draft), 1973.

Carter, C. *Handbook of mental retardation syndromes.* (2nd ed.) Springfield, Ill.: Charles C. Thomas, 1970.

Combs, R. and Harper, J. Effects of Labels on attitudes of educators toward handicapped children. *Exceptional Children*, 1967, 33. 399-403.

Dunn, L. Special education for the mildly retarded - is much of it justifiable? *Exceptional Children*, 1969, 35, 5-22.

Fischer, C. Intelligence defined as effectiveness of approaches. *Journal of Consulting and Clinical Psychology*, 1969, 35, 5-22.

Fischer, C. The testee as co-evaluator. *Journal of Counseling Psychology*. 1970, 17, 70-76.

Fischer, C. Contextual approach to assessment. *Community Mental Health Journal*, 1973, 9, 38-45.

Hora, T. Cognition and health. *Journal of Religion and Health*, 1962, 1, 283-291.

Jaslow, R. and Smith, S. A proposal for a new conceptual use of the term "mental retardation." *Mental Retardation*, 1972, 10, 36-37.

Jones, R. Labels and stigma in special education. *Exceptional Children*, 1972, 38, 553-564.

Mercer, J. I.Q.: The lethal label. *Psychology Today*, 1972, 6, 44-47.

Meyerowitz, J. Self-derogations in young retardates and special class placement. *Child Development*, 1962, 33, 443-451.

88X

ser, C. General principles of diagnosis and management of mental retardation. In C. Poser (ed.) *Mental retardation diagnosis and treatment.* New York, Evanston, and London: Harper & Row, 1969.
ecalfa, L. Self-determination and the concept of man. In paper presented at annual meeting, American Psychological Association, Washington, D.C., September, 1971.
ecalfa, L. Is secrecy in the client's best interest? *Journal of Rehabilitation*, 1972, 38, 19-20.
ecalfa, L. Rapport or manipulation? *Journal of Rehabilitation*, 1973, 39, 14-15. (a)
ecalfa, L. Some uses and abuses of client records. Paper presented at annual meeting, American Psychological Association, Montreal, Canada, August, 1973. (b)
Sarason, S. and Doris, J. *Psychological problems in mental deficiency* (fourth ed.) New York, Evanston, and London: Harper and Row, 1969.
Sterewu, L. Doctor-patient relationships: New ways in an old system. Paper presented at annual meeting, American Psychological Association, Montreal, Canada, August, 1973.
Thomas, A., Hertzig, M., Dreyman, I., and Fernandez, P. Examiner effect in IQ testing of Puerto Rican working-class children. *American Journal of Orthopsychiatry*, 1971, 41, 809-821.
Vergason, G. Accountability in special education. *Exceptional Children*, 1973, 39, 367-373.
Watson, P. IQ: The racial gap. *Psychology Today*, 1972, 6, 48-50, 97, 99.

Mary Ann Thomas received an M.A. in religion from Northwestern University and an MSW from the University of Chicago. She is a Psychiatric Social Worker at the Gaston County Mental Health Center and an Instructor in Social Work at Sacred Heart College in Charlotte, North Carolina.

POETIC PERSPECTIVES OF A TODDLER

MARY ANN THOMAS

AT RAIN

Why does the thunder bark?
Why does the thunder?

It will not hurt me;
the thunder and lightning
will shake hands.

Sing me a song, mommy,
the one about the star wandering
and twinkling.

Sing it too-gain.

BODY BALLOONS

I see me free
as a balloon, floating high up up,
red now, yellow tomorrow,
always filled, round, round
round the room, free
a leak, a scratch,
and pffft–

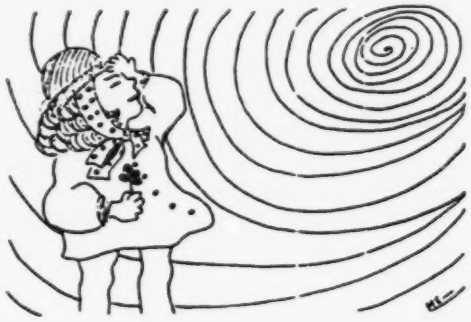
I'm angry with Mommy
I'm going to let out all of her air.

T SUN

ll that sun to go back up
in the sky,
t at to stop pinching my eyes.

es it hurt the animals, the
elphanant, like me?
can they put their eyes in
their pockets?

I put my eyes in my pockets.
en the sun can
n on me all night.



ODE TO A TOILET

Monster: large as an elphenant
hard and cold like wet stones
with your funny white hat
and the biggest mouth I've ever seen

Do you like the paper I feed you?
You must. You take everything
and swish it away
with one wink of your silver eye.

Sometimes you get stomach aches
and Mommy makes you burp.
Do you like your blue messidin?
Such a fuss when I give you my dodo

and even Daddy says
he's prout of me.
But please
don't swallow me too.

Illustrated by Patricia McGlothlan

BEST COPY AVAILABLE

Journal of Clinical Child Psychology, 1976, Vol 5, No. 1

Louise Mead Riscalla is a Clinical Psychologist at the Woodbridge Emergency Reception and Child Diagnostic Center. She was formerly with the New Jersey State Diagnostic Center where her work primarily included court referred diagnostic assessment of children. Her articles on attitudes of professionals toward "patients", rehabilitation, psychological assessment, chronic illness, and psychosomatic medicine have appeared in numerous journals. She is currently involved in issues concerning the mis-diagnosis and labeling of brain-damaged children, and has done research in neurophysiology and neuropsychology which has been published in medical journals.

THE PROFESSIONAL'S ROLE AND PERSPECTIVES ON CHILD ABUSE

Louise Mead Riscalla

There is a tendency to consider the eighth amendment pertaining to cruel and unusual punishment as relating to forms of corporal punishment inflicted on children by adults, such as parents or guardians, institutional staff, school personnel, et cetera. While professionals generally desire to help children, professionals inadvertently abuse children in ways which tend to be overlooked at the expense of the more extreme forms. Some of the ways in which professionals inadvertently or deliberately abuse children will be explored. While much of the material can be applied in principle to all individuals working with children, the emphasis will be in the educational area because most of the waking hours of children are devoted to school or school related activities.

The Children's Defense Fund (1974) did an extensive study on children who are not in school. Of particular relevance to child abuse, are children who are harassed and rejected by school personnel to the extent that education becomes so intolerable that they may react by truanting or dropping out. These rejected children include, for example, migrant children, pregnant girls or unwed mothers, children labeled as "troublemakers" by some teachers, those not expected to achieve academically and as a consequence are not given attention or helped to learn, children who are so poor that they have few clothes and no money for school activities, and members of minority groups in desegregated schools who are more or less pushed out (Children's Defense Fund, 1974). It may be observed that children and their parents have been encouraged to withdraw through counselling, and as a consequence, are pushed out of school through a counselling process designed to "help" children. Psychologists are frequently placed in a conflict between the best interests of the child and an educational system, including administrative policies of the facility which often take priority.

There are instances where it is necessary to seek expert help from outside sources for children in the school. However, there is a risk that psychologically sophisticated teachers and school psychologists could seek expert help from outside sources without considering the possibility of what Mays referred to as "pseudo-maladjustment" (Mays, 1973, p. 45) where a child avoids issues so as not to face the challenge of demands placed on him for more difficult assignments which give him less time for recreation. As a consequence, teachers and parents might encourage this "pseudo-maladjustment" by trying to treat it. It is important to recognize that at some point, practically all children reflect textbook symptoms of psychological disturbances without being disturbed enough to necessitate outside help.

The pressures of teachers and parents obsessed by marks or psychologists by I.Q. scores and achievement tests often produce irritability, fatigue, psychosomatic disorders, and rivalry between classmates. "Many schools divide the students into separate groups based on ability or post-graduate career plans, thereby setting up social-class rivalries based on rank and status" (Haney and Zimbardo, 1975, p. 106). Competition is often encouraged because it helps children increase their school performance. However, in competition, one's victory is at the expense of another's defeat; often accompanied by humiliation and a fear of failure. As a consequence, students are more concerned with competing for grades than with the subject matter to be learned and go through school without really learning. The mental health of those working with children is important and emotional problems can have harmful consequences on the student. For example, power hungry teachers belittle their students, sadistic teachers are physically and verbally cruel, and some teachers unconsciously manipulate students to serve their own needs, et cetera. It was reported that a teacher in Columbus, Ohio was so sarcastic and hostile to a student that the girl faked stomach aches to stay home, fell behind in her work, and then refused to go to school (Brenton, 1971). It is difficult to deal decisively with teachers, school psychologists, and other school personnel largely because of tenure protecting them and organizations who fight for the rights of professionals. There is also a tendency for school administrators and public officials to assume a "don't rock the boat" attitude with the consequence of encouraging and perpetuating child abuse. An alternative perspective based on enlightenment recognizes that the health of the professional is basic to professional effectiveness, and is more important than technique — which is ordinarily stressed. It is important for those working with children to be aware of their own motivations and "hang-ups," and to be open to opportunities for growth as individuals.

Child abuse is an extreme form of discipline, but in the final analysis, discipline is discipline regardless of degree. For purposes of maintaining the social structure, including the institutions of society, various forms of discipline are accepted and practiced. Discipline policies in school are arbitrary and not

90x

many school districts have clear, written policies governing their disciplinary actions. A number of incidents have been reported where corporal punishment is still being used (Maurer Ed., 1975). New Jersey and Massachusetts are the only states which have laws prohibiting corporal punishment (Haney and Zimbardo, 1975). Some schools give students a choice between suspension, corporal punishment, or detention (Children's Defense Fund, 1974). Discipline is one of the manifestations of a desire to be psychologically secure. Since discipline guarantees a result, the desired result is more important than the means used to attain it. In the use of discipline to obtain a particular result, the system often becomes more important than the human beings in it and discipline then becomes a substitute for or an erroneous conception of love. Rewards and punishment are aspects of discipline and part of tradiltional child rearing practices. The individual is bribed into "right" action by rewards and is instilled with fear of and actual punishment if violations occur. The reward or punishment then often becomes more important than actions. A perspective involving rewards and punishment assumes that such a system of discipline is a means to an end whereby the individual will eventually "act right" for its own value without any ulterior motivations or expectations. This is difficult and practically impossible, particularly in view of the process of conditioning. An alternative perspective does not seek immediate results, but is primarily concerned with explaining and encouraging consideration for others including harmonious action without inducements or threats. It is recognized that the means are more important in determining the result, rather than the result being more important than the means in the more or less traditional perspectives.

Children are often inadvertently abused through programs designed to help them. The general field of special education including special services was developed to provide opportunities for those children who deviated from "normal" children to help them attain their maximum potential. The perspective of special class placement is based on the traditional medical model which assumes that by finding out what is "wrong" with an individual, corrective measures including treatment can then be undertaken. Children are placed in special education by finding out what is "wrong" through diagnosis and treatment in the form of remediation by special class placement or special services. The deleterious consequences of labeling children in special education is well known and has been dealt with extensively (i.e. Combs and Harper, 1967; Dunn, 1969; Jones, 1972; Macmillan, Jones, and Aloia, 1974; Mercer, 1972; Riscalla, 1974). Ideally one objective in special education is to eventually return the child to the regular classroom setting. However, too often "special education children" remain in special class placement through their school careers.

An enlightened perspective is concerned primarily with the uniqueness of each individual rather than the majority of students; and recognizes that not every child can benefit from traditional sitting-in-class being taught by a teacher. Children are described as individuals and understood in the context of their own lived, here and now world. There is evidence of enlightened school programs in school systems throughout the country. A number of school districts have developed new schools or programs to serve as alternatives for those children who do not respond to traditional methods of education. These alternatives include for example, work-study programs where students in vocational programs can earn some money while completing their education and have the possibility of employment following graduation. The Metropolitan Youth Education Center in Denver has an alternative program for senior high school students and classes are held all year, both day and evening so that students can have a flexible schedule (cited in Children's Defense Fund, 1974). Richland County School Number One in South Carolina has been operating a Walk-In School since September, 1972 and has no failures or formal schedule (cited in Children's Defense Fund, 1974). Special education in most instances by definition and philosophy can be carried out in a regular classroom (Siegal, 1969). Severely handicapped children, such as those with epilepsy (Tenny and Lennox, 1962), blind (Gray, 1962), and mentally retarded (Blackman and Sparks, 1965) have been intentionally enrolled in classes with nonhandicapped children. It has been mandated that handicapped children participate in Head Start (Cohen, 1975), thus integrating handicapped with nonhandicapped children. Courts are requiring "least restrictive" placements of handicapped children where the primary concern is to place a child in the most normal setting (Russo, 1974).

Children's rights legislation protecting the rights of children can have harmful consequences, particularly if the legislation is carried to extreme forms. Teachers or school administrators have had judicial, executive, and legislative powers. Haney and Zimbardo (1975) reported similarities between high schools and prisons. "As in the country's prisons, America's schools give their guards or teachers almost absolute authority over the student inmates" (Haney and Zimbardo, 1975, p. 26). Children are now entitled to due process in the school system. The complexity and length of the procedure of due process present difficulties in application to a school setting and if "applied rigorously in school the educational organization would probably come to a halt" (Duffee, 1974, p. 57). With suspensions, the Supreme Court only provided "minimum" due process, as the student before suspension is not consitutionally entitled to have a lawyer, cannot call on his own witnesses to testify, or cross examine witnesses. Due process poses a dilemna in that suspensions are often given to discipline unruly students in order to protect the rights of other students and school property, and at the same time to protect the rights of each individual child. There is a problem in balancing the rights of the states to have meaningful, peaceful schools and concurrently in protecting the constitutional rights of children. "For all its vagueness, disjointedness and seeming contradictions, the 'minimum' due process requirement for students which the Supreme Court now has laid down in Goss vs. Lopez must somehow be made to work. If this turns out to be impossible (and the odds seem omniously in favor of such an outcome) the cure perscribed in a future ruling is likely to be more painful than the disease" (Nolte, 1975, p. 49). The rights of children are often at the expense of the rights of parents, teachers, and institutions which generate a considerable amount of resistence and ways of finding legislative loopholes. Children with little or no respect for others are often protected by the law and manipulate over zealous child advocates. It is possible that this situation could perpetuate delinquency. It often happens that enlightened parents, teachers, and other professionals concerned with children are abused in the process.

The courts have acted to remove children who have been neglected or abused from their homes. However, the child frequently "goes from the frying pan into the fire" by being

BEST COPY AVAILABLE

placed in inadequate foster homes where children encounter neglect, abuse, or overcrowded institutional facilities including group homes. For example, both of Karen's parents were alcoholic and she was abused and neglected to such an extent that she was placed in a number of foster homes. In one of the foster homes she became pregnant by her appointed guardian necessitating removal from the home and placement in a diagnostic center for evaluation and recommendation for placement. Karen maintained that she wanted to be reunited with her mother in spite of the fact that her mother had previously abused and neglected her. It may be observed that children often perceive removal from the home as a further punishment, and have difficulty understanding the real reason for their removal. They also display a considerable amount of loyalty and devotion to their parents, regardless of their mistreatment, and have the belief that it is they who are "bad." "Efforts made to 'save' the child from his bad surroundings and to give him new standards are commonly of no avail, since it is to his own parents who, for good or ill, he values and with whom he is identified" (Bowlby, 1965, p. 80). The human element can be disabling and interfere with decisions rendered in behalf of children, including the judicial process. For example, Burt (1972) indicated that the desire of the judge to "punish the parents and remove their child when he believes that they have acted (and in fact they have acted) hurtfully toward the child, can be as much or more a product of the judge's unconscious identification in the transaction as his reasoned response that this individual child will best be helped by removal from these particular parents" (Burt, 1972, p. 98). Children's rights are determined in the adversary system of enforcing some of the rights of parents engaged in conflict which treats children as property (Freed, 1972) and can, in principle, pertain to all instances of child abuse and neglect.

There are circumstances where a parent takes a child to court for incorrigibility or files a complaint as a desperate, last resort measure to seek and obtain help for the child. The child is now entitled to due process including representation by an attorney. As a consequence, the adversary position of parents and children is increased which often results in a further breakdown of family relationships where everyone, including the child is abused.

The vagueness of much of the legislation and difficulty in application or enforcement in the final analysis is due to a belief in working individually or together toward some ideal or future goal. Individuals are then shaped according to beliefs in what should be; and as a consequence, the ideal or future goal becomes more important than the individual in the context of his lived in world with all of his complexities. There is also a gap between the present and future in which many factors exert an influence and are often ignored or overlooked by a primary focus on endeavoring to bring about what should be. The alternative is a broader perspective based on understanding an individual or set of circumstances in his or her present context without preconceived ideas as to the "should be" ideals and goals; and to respond on the basis of what is. (Freed , 1972) and Foster (cited in Freed, 1972) in the area of family law have been instrumental in encouraging an alternative perspective of children with the moral and legal right to be considered as persons. "As persons, children have individual interests apart from — and sometimes in conflict with-parental or societal interests. Children are entitled to assert these interests and to have them heard and considered by any authority rendering a decision on them" (Free, 1972, p. 36). Pertinent evidence is frequently not permitted, overlooked, or not included for purposes of "winning a case" or to gain some advantage in an adversary system. From the alternative perspective, Freed (1972) advised an approach where " . . . it is the duty of the court to admit all relevant evidence bearing upon the actual psychological and physical welfare of the child and to base the decision upon such evidence. Independent counsel for the child should be permitted to participate in the hearing with the same rights as counsel for the parents" (Freed, 1972, p. 38). As a consequence, the child is perceived as a unique person and recognized in his own rights.

Institutions such as schools, the legal system, government, etc. are, in the final analysis, composed of people and exert a powerful influence over the lives of others. Individuals with an enlightened perspective have the constructively critical skills and values necessary to resist and overcome with positive alternatives, the manipulation and dehumanization that takes place within the institutions of society. It is therefore hoped that in the process of evolution without revolution, there will be no need to legislate or specify that children are persons, for it will be assumed as a matter of fact. What is best for the child will then take place as a means to an end based on a perspective of children as persons, and will therefore be beneficial. "People who are awake to realities, aware of their inner selves, and accurately informed of their environment can best contribute to the social, moral, and political reforms vital for an enlightened society" (Riscalla, 1971, p. 131).

Woodbridge Emergency Reception and Diagnostic Center
P.O. Box 203
Avenel, New Jersey 07001

92X

The author is a Clinical Psychologist at the Woodbridge Emergency Reception and Child Diagnostic Center. Much of the material in this article is based on her previous position at the New Jersey State Diagnostic Center, which primarily included court referred diagnostic assessment of children. Her articles on attitudes of professionals toward "patients," rehabilitation, psychological assessment, children's rights, chronic illness, and psychosomatic medicine have appeared in numerous journals.

ASSESSMENT AND CLASSIFICATION: ACCENT THE POSITIVE, ELIMINATE NEGATIVES

Louise Mead Riscalla

ABSTRACT

The implications of some legislation and professional standards pertaining to the assessment and classification of children with special regard for their individual rights will be discussed. Ways of working within the establishment, while using legislation, professional standards, and innovative assessment methods will be illustrated with personal anecdotes and case material. Since there are so many tests, varied settings and populations tested, this article will be based on personal experiences as a clinician in my former position at a state diagnostic center which utilized a team approach. While the emphasis is on children's rights, much of the material to be discussed could, in principle, be applied to all aspects of testing and individual rights.

The clinician must often face the dilemma created where an individual has the right to refuse a psychological assessment, yet the clinician is required to perform this assessment as part of an administrative or legal procedure. I have worked with a number of children who have refused assessment, and generally I quite frankly and openly admit the conflict between respecting their rights and doing what is required of my position. Their reasons for refusal are also explored. Many children have related that they have been tested before so many times that they just cannot see the reasons why they have to be tested again and do not believe that repeated testing is necessary. The child is then given an explanation including reasons as to why it is necessary to repeat the tests as well as why one should cooperate in the testing procedures. When I have seen that previous test scores were lower than the child's observed performance or than I have suspected the child to be, I will usually tell a child that the tests before did not come out very well and I think he could do better if he really tried. Very often a child will admit that he took previous tests not caring or doing his best, and with encouragement, willingly enters testing. If a child still objects or refuses to be tested after an explanation, the child is then told to advise when testing is desired. This procedure provides the individual with a sense of responsibility by deciding when the tests will be taken. The feelings and rights of the child are respected by testing when the child feels ready and is motivated. It is important to make time available for giving children the tests when they ask for them. If this time is not available, it is often expedient to explain the reasons to the child and then mutually decide on a time for testing. A more accurate assessment is likely to be made by testing a child on the child's terms rather than making the child conform to a pre-set testing schedule. When a child outrightly refuses testing, the report includes documentation of the evidence regarding reasons for the child's refusal such as the child's behavior during testing, attitudes toward the tests or examiner, attempts to explain reasons for testing, etc. The child is usually told what is said in the report and why, because of the individual's right to know. If a child shows strong evidence that he is not interested in knowing or does not care, his desire *not to know* is also respected.

The individual's right to privacy is an issue that is often encountered. It may be observed that in many ways the field of psychological testing has been considered by some as a form of voyeurism used under the pretence of helping people. While trying to gain information in order to give a fair evaluation, it is often necessary to become involved in very personal issues. When children do not want to discuss specific highly personal information necessary for testing or even to relate any matters of a personal nature, there is no hesitation about apologizing for having to ask personal questions. An explanation is generally given as to why such personal questions are asked.

It has been recognized that the course of the lives of individuals, especially children, can be greatly influenced by assessment and classification, particularly in the case of standardized intelligence testing. Since human potentials are infinite, tests and classification procedures place limitations on an individual which can be harmful. It is essential to recognize the limitations of tests and classification procedures so that individuals can be free to develop their inherent potentials. The Manual for the *Wechsler Intelligence Scale for Children-Revised* clearly states that the intention of the test battery "is to assess a child's performance under a *fixed set of conditions*, and not to test the limits of his knowledge (Wechsler, 1974, p. 43)." It is important to specify in the report what the intelligence test is, i.e., a set of conditions. The administrative procedures of a facility often require a psychologist to specify the IQ score and to include all of the scaled sub-test scores in the report. However, the *Standards for Educational and Psychological Tests* (APA, 1974) indicate that "users should avoid the use of terms such as IQ, IQ equivalent, or grade equivalent, when other terms provide more meaningful interpretations (p. 70)." Consequently, I have found myself in a dilemma between professional standards and administrative procedures. I have resolved this dilemma in a number of ways, many of which have been reported in the literature (e.g. Fischer, 1969, 1971; Mercer, 1972; Riscalla, 1972, 1974b). My interpretation of the IQ is based on all of the tests administered, observation of the child, conversations with him, and the available history. For example, when I have tested a child and have reasons to believe that he is intellectually higher than reflected by test

93X
Vol. 5, No. 3
109

scores, I will use terms such as "ordinarily placing him at," then indicate the range of intelligence in the test manual, and document from evidence on hand where I believe he is and why he did not do as well on present testing. I will also use a phrase such as "numerically placi[illegible]" then specify the range of intelligence indicated in the [illegible] manual. As observed in the literature, there is no defi[illegible] conclusion as to what intelligence is, nor how socio-cultural factors, economic influences, motivation of the child, personal qualities of the examiner, and a multitude of other factors, singly or in combination, influence IQ scores. A general and tentative interpretation of the IQ helps to protect the rights of the children who bear the consequences of testing and is consistent with the test standards (APA, 1974) and manual (Wechsler, 1974). The child is generally informed of the limitations of the test and given the range when the child wants to know, but the numerical scores are not disclosed with the explanation that a numerical score depends on the particular test used. Children who insist on knowing generally accept this explanation. The child is also informed of the factors which influence test scores (and are indicated in the written report). Recognizing that test scores can vary for a multitude of reasons, phrases such as, "at this time," or "on present testing" are often included in a report to indicate that the individual is constantly changing and never static. In the final analysis, there is no conflict over the IQ when one is adhering to the test manual and standards and being open with the child.

Classification and diagnosis are based on a medical model which assumes that by finding out what is wrong with an individual, a course of treatment can then be undertaken (Riscalla, 1976). The harmful consequences of labeling children, particularly in special education, are well known and documented in the literature (e.g. Combs and Harper, 1967; Dunn, 1969; Macmillan, Jones, and Aloia, 1974; Mercer, 1972; Riscalla, 1974a). While classification and diagnosis are still considered necessary for purposes such as funding, keeping statistical records, placement and treatment considerations, etc., it can be done in a constructive manner. For example, I often include documented evidence that a child is reacting to social, racial, economic, educational, or family problems. A description of the child as an individual and his needs could be a practical, viable alternative to classification and should be used whenever possible. Every person and life experience is unique. A technique used to administer, score, and interpret a test and write a report which is satisfactory for one individual may not be for another. A technique is the by-product of the human qualities and perspective of the test maker, administrator of the tests, and the individual taking the tests. It is well known and frequently stated in test manuals, literature, standards, and by those working in the field, that the instrument can only be as valid as those who make the tests and those who use them. Thus, in effect, the test maker and user become the assessing instrument. In the final analysis, the personal qualities and perspective of the professional are of tantamount importance in assessment and classification procedures.

REFERENCES

American Psychological Association. *Standards for educational and psychological tests*. Washington: American Psychological Association, 1974.

Combs, E. and Harper, J. Effects of labels on attitudes of educators toward handicapped children. *Exceptional Children*, 1967, 33, 399-403.

Dunn, L. Special education for the mildly retarded — is much of it justifiable? *Exceptional Children*, 1969, 35, 5-22.

Fischer, C. Intelligence defined as effectiveness of approaches. *Journal of Consulting and Clinical Psychology*, 1969, 33, 669-674.

Fischer, C. Toward human research via a moratorium on IQ-testing. *Clinical Child Psychology Newsletter*, 1971, 10, 4.

Macmillan, D., Jones, R., and Aloia, C. The mentally retarded label: A theoretical analysis and review of research. *American Journal of Mental Deficiency*, 1974, 79, 241-261.

Mercer, J. IQ: The lethal label. *Psychology Today*, 1972, 6, 44-47, 95-97.

Riscalla, L. The captive psychologist and the captive patient. *Professional Psychology*, 1972, 3, 375-379.

Riscalla, L. Mental retardation: Fact or conjecture? *Journal of Clinical Child Psychology*, 1974, 3, 43-45. (a)

Riscalla, L. Records: A legal responsibility. *Journal of Rehabilitation*, 1974, 40, 12-14 (b)

Riscalla, L. The professional's role and perspectives on child abuse. *Journal of Clinical Child Psychology*, 1976, 5, 24-26.

Wechsler, David. The manual for the Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation, 1974.

Woodbridge Diagnostic Center,
P.O. Box 203,
Avenel, New Jersey, 07001

Developmental Arts for Exceptional Children: Program Report

Judy Tillinghast

The author is a music therapist/teacher who obtained her B.M.E. in Music Education from Oklahoma City University. She is currently obtaining her masters degree in Institutional Counseling and Music Therapy under the Master of Arts in teaching program at OCU.

A developmental arts (DA) therapy program was developed and implemented at the Child Study Center, Oklahoma Children's Memorial Hospital, as a trial program to assess, on an observational basis, the efficacy of such a program. Based on the results a developmental arts program appears to be a useful and integral part of a child's therapeutic schedule. The purpose of a DA program is sensory stimulation with much focus on perceptual-motor activities. The sensory modality concentrations are visual, auditory, and touch. Adoption of the title Developmental Arts Therapy for the program is appropriate due to the intentional combination and independent utilization of music, art, and movement as therapeutic vehicles.

The goal of the DA Therapy program, whether individual, large or small group therapy, is to stimulate the child through the medium of the arts, reinforce the child's response, and activate the child's initiative to respond independently. A natural advantage of the arts as a vehicle is the instinct to respond to rhythmical stimulation. Helpful too is a child's love of music and the child's fascination of various textures (rough, smooth, soft, etc.). In an effort to categorize the goals, the concentration is presented as perceptual-motor, and sensory integrative. The latter is defined as the process of seeing, hearing, and feeling the content of the activity, whether externally manipulated or achieved by the child independently.

ASSESSMENT AND CLASSIFICATION: ACCENT THE POSITIVE, ELIMINATE NEGATIVES*

Louise Mead Riscalla

New Jersey State Diagnostic Center

Edison, New Jersey 08837

It may be observed that there has been a considerable amount of controversy in the area of psychological assessment. For example, there has been litigation concerning the mislabeling of individuals on the basis of assessments (examples cited in Harris, 1972; Vergason, 1973) and claims of racial and cultural biases especially with intelligence and achievement testing (e.g. Baezer, 1973; Frank and Kagan, 1973; Holmen and Docter, 1974; Mercer, 1972; Watson, 1972; Williams, 1970). The Standards for educational and psychological tests, prepared by a joint committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education were revised (APA, 1974). Among other things, this revision of the test standards cautions test users about cultural and personal biases, warns against labeling, indicates that it is essential that the test manual warn against common misuses of tests, and includes the right of the individual tested, his agent, or guardian the right to know his score and interpretation. The Manual for the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) recognizes that the diagnostic skill of the psychologist depends on the ability to interpret and detect unusual and abnormal responses; and an awareness of the extent of the influence of socio-economic and cultural background on a subject's responses.

Although many individuals would like to abandon tests, it would be impossible to abandon the use of psychological tests, because the educational system, employment practices, institutions, government agencies, etc. require some criteria for classification and remediation purposes. In addition to the fact that an abolishment of psychological testing could, for example, result in massive unemployment for the testing industry as well as other disciplines using tests, the positive aspects of services rendered would be lost.

This presentation will discuss the implications of some legislation and professional standards pertaining to the assessment and classification of children with special regard for their individual rights. Ways of working within the establishment, while utilizing legislation, professional standards, and innovative assessment methods will be illustrated with personal anecdotes and case material. Since there are so many tests, varied settings and uses, and populations tested, the presentation will be based on personal experiences as a clinician

---

*Part of a symposium, "Psycho-educational classification and public policy: Children's rights, presented at the annual meeting, American Psychological Association, New Orleans, August, 1974. Gerald P. Koocher, Chairman.

working in a state diagnostic center using the team approach. The primary function of the clinician is the psychological assessment of individuals referred by courts throughout the state for various offenses (e.g. breaking and entry, homicide, armed robbery, vandalism, etc.). My particular assignment consists of assessing children on the in-patient service who range in age from eight to 18 of both sexes. These children come from practically all cultural and socio-economic backgrounds. Many have a history of school problems such as truancy, special class placement, underachievement, etc.. A number of children have been abused by their families, and others have experienced a split in the family by divorce, seperation, or death of a parent. While the psychologist is relatively free to select the tests he uses in a battery, the basic test battery generally consists of an individual intelligence test such as the Wechsler Intelligence Scale for Children-Revised, Wechsler Adult Intelligence Scale, Rorschach, Thematic Apperception Test, House-Tree-Person, Chromatic House-Tree-Person, Bender Gestalt, and a Sentence Completion Test with questionairre specifically designed for use at the diagnostic center. Therefore, it is believed that while the emphasis is on children's rights, much of the material discussed in this presentation could, in principle, be applied to all aspects of testing and individual rights.

The clinician must often face the dilemma created by an individual's right to refuse being tested on one hand, while being required to perform such an assessment as part of an administrative or legal procedure on the other. I have encountered a number of children who have refused assessment, and generally quite frankly and openly admit the conflict I have between respecting their rights and doing what is required by the courts and my position. I also explore their reasons for refusal. For example, many children have told me that they have been tested before so many times that they just can't see the reasons why they have to go through it again and don't believe that repeated testing is necessary. In such situations, I usually tell the child that I understand how he feels and that I would feel the same way if in his position. However, since psychological testing is one of the reasons for sending a child to the diagnostic center, I also tell this to the child and ask for his help in completing the necessary tests. When I've seen that previous test scores were lower than I've observed or suspected the child to be, I'll tell a child that the tests he had before didn't come out so well and I think he could do better if he really tried. Very often a child will admit that he took previous tests not caring or doing his best, and with encouragement, willingly enters testing trying to do his best largely because of the respect shown for his feelings and rights. When I've seen a child refusing to take tests after several attempts to test him, I'll frankly tell the child that although I understand his predicament, I would like to do the testing and to let me know when he "feels up to it." Such children have approached me when ready, and I always make certain that time is available to give these children the tests when they ask for them. When a child outrightly refuses, I describe the child's behavior, my attempts to discuss testing with the child, and endeavor to document this refusal in the report. I usually tell such a child what I am saying and why because of the individual's right to

know. If a child shows strong evidence that he is not interested in knowing or doesn't care, his desire not to know is also respected.

The individual's right to privacy is an issue that confronts me daily in my work. I have observed that in many ways the field of psychological testing has been considered by some as a form of voyeurism used under the pretence of helping people. While trying to gain information in order to give a fair evaluation, it is often necessary to become involved in very personal issues. I usually tell the children exactly how I feel about it, especially when they do not want to disclose particular material or tell me some things of a personal nature. Under such circumstances, I have no hesitation about apologizing to a child for having to ask him some personal questions and explaining why I'm asking such questions. I also admit my helplessness and frustration with the "system."

It has been recognized that the course of the lives of individuals, especially children can be greatly influenced by assessment and classification, particularly in the case of standardized intelligence testing. Since human potentials are infinite, tests and classification procedures place limitations on an individual which can be harmful in many ways. It is essential to recognize the limitations of tests and classification procedures so that individuals can be free to develop their inherent potentials. The Manual for the Wechsler intelligence scale for children-revised clearly states that the intention of the test battery "is to assess a child's performance under a fixed set of conditions, and not to test the limits of his knowledge (Wechsler, 1974, p. 43)." It is then important to specify in the report what the intelligence test is-a set of conditions. In my work it has been and is the practice to specify the IQ scores, include all of the scaled sub-test scores when placement for some facilities are recommended, and to specify scaled sub-test scores as they pertain to test interpretation in the report. It is necessary to give the quantitative scores for diagnostic and classification purposes in making a recommendation for the child. The Standards for educational and psychological tests (APA, 1974) indicate that "users should avoid the use of terms such as IQ, IQ equivalent, or grade equivalent when other terms provide more meaningful interpretations (p. 70)." Consequently I am often placed in a dilemma between professional standards and doing part of my job. I've resolved this dilemma in a number of ways, many of which have been reported in the literature (e.g. Fischer, 1969, 1971; Mercer, 1972; Riscalla, 1972, 1974b). My use of the IQ involves all of the tests administered, observation of the child in his daily activity at the center and in my office, conversations with him, and the available history. For example, when I've tested a child and have had reasons to believe that he is intellectually higher than reflected by test scores, I'll use terms such as "ordinarily placing him at" indicating the range of intelligence in the test manual, and then document from the evidence on hand where I believe he is and why he didn't do as well on present testing. I'll also use a phrase such as "numerically placing him" then specify the range of intelligence indicated in the test manual. I am deliberately general in my interpretation and reports dealing with the IQ because of the controversy in the field which, as may be observed in the literature, has not reached

BEST COPY AVAILABLE

any definite conclusion as to what intelligence is, the socio-cultural factors, economic influences, motivation of the child, personal qualities of the examiner, and multitude of other factors which singly, or in combination, influence IQ scores. I am also general and tentative in dealing with the IQ in order to protect the rights of the children who bear the consequences of such testing and to adhere to test standards and manual which indicate that the IQ can vary with a number of factors (e.g. APA, 1974, Wechsler, 1974). I generally inform the child of the limitations of the test, provide him with the range when he wants to know, but do not give the numerical scores with the explanation that a numerical score depends on the particular test used. Children who insist on knowing generally accept this. The child is also informed of the factors which influence his test scores. These factors are also contained in the written report. Recognizing that the test scores can vary for a multitude of reasons, I usually include phrases in the report such as, "at this time," "on present testing," etc. indicating that the individual is constantly changing and never static. I believe that in the final analysis, there is no conflict over the IQ when one is adhering to the test manual, standards, and being open with the child.

Diagnosis is another problem, and again, I am deliberately general because I do not know all of the answers. For example, I use terms such as "probably, "could be," "maybe," "it is conceivable," etc. in the report because I think that the use of such terms is "telling it like it is." I describe what I see of an individual and usually conclude my report, for example, by "over-all test evidence, observations, discussions, and history give the impression of an individual whose problems are reactive in essence." If a child is overtly psychotic, I describe the child in the context of testing, but while coming close, do not actually pin the label on him by use of a phrase such as "psychotic-like." I have no hesitation when the evidence is present and document it, by indicating that a child is reacting to a racial, economic, or family problem. When I'm not sure of some of my interpretations, I specify these uncertainties in the report and often check my interpretations with the child. A contextual approach to testing (e.g. Fischer, 1973; Riscalla, 1972) is extremely important, and this contextualizing is consistent with the test manuals, standards, and could in many instances hold up in court.

There has been concern with the right of the individual or his representative to review records, including the reports of psychological assessments. For example, among other things, the Guidelines for the collection, maintenance, and dissemination of pupil records (Russell Sage Foundation, 1970) recommended procedures where the student or his parents could challenge any information contained in the student's school records. New York State Education Commissioner Ewald Nyquist ruled that parents can see all school records and information regarding their children (Buder, 1972). According to a New Jersey Supreme Court Decision, (State of New Jersey, 1969), the patient or his representative has the right to subpoena the report of the psychiatric evaluation, including the report of the psychologist, and to contest the findings. In my many years of experience I have not as yet been subpoenaed, which I hope

98X

means that I am respecting the rights of the children and at the same time fulfilling the requirements of my position.

While there are standardized procedures, techniques alone are meaningless and inauthentic (Riscalla, 1973, 1974a). Every person and life experience is unique. Therefore, a technique used to administer, score, interpret a test, and write a report which is satisfactory for one individual may not be for another. A technique is the by-product of the human qualities and perspective of the test maker, administrator of tests, and the individual taking the tests. It is well known and as observed, has been frequently repeated in test manuals, literature, standards, and those working in the field, that the instrument can only be as valid as those who make up the tests and those who use them. Thus, in effect, the test maker and user become the assessing instrument. While legislation, professional standards regarding tests, court action, etc. are helpful, in the final analysis, as implied by test standards and manuals (e.g. APA, 1974; Wechsler, 1974), the personal qualities and perspective of the professional are of tantamount importance in respecting children's rights.

References

American Psychological Association. Standards for educational and psychological tests. Washington: American Psychological Association, Inc., 1974.

Beezer, B. Testing: Educational goals and change. Intellect, 1973, 102, 114-115.

Budner, L. Nyquist rules parents may see children's full school records. New York Times, 1972, 121, 1, 30.

Fischer, C. Intelligence defined as effectiveness of approaches. Journal of Consulting and Clinical Psychology, 1969, 33, 668-674.

Fischer, C. Toward human research via a moratorium on IQ-testing. Clinical Child Psychology Newsletter, 1971, 10, 4.

Fischer, C. Contextual approach to assessment. Community Mental Health Journal, 1973, 9, 38-45.

Harris, T. An introduction I.Q. abuse. Psychology Today, 1972, 6, 39.

Holmen, M. and Docter, R. Criticisms of standardized testing. Today's Education, 1974, 63, 50-60.

Mercer, J. IQ: The lethal label. Psychology Today, 1972, 6, 44-47, 95-97.

Riscalla, L. The captive psychologist and the captive patient. Professional Psychology, 1972, 3, 375-379.

Riscalla, L. Rapport or manipulation? Journal of Rehabilitation, 1973, 39, 14-15.

Riscalla, L. Mental retardation: Fact or conjecture? Journal of Clinical Child Psychology, 1974, 3, 43-45. (a)

Riscalla, L. Records: A legal responsibility. Journal of Rehabilitation, 1974, 40, 12-14. (b)

Russell Sage Foundation. Guidelines for the collection, maintenance, and dissemination of pupil records. Russell Sage Foundation, Hartford, 1970.

State of New Jersey v. Thomas Horne, A-37; Mailyor Andrews, A-145; Booker T. Blanford, A-146; James Coleman, A-147; Norris Barnes, A-149. Supreme Court of New Jersey, September, 1969 term.

100 X

...lla

Vergason, G. Accountability in special educa... ...ceptional Children, 1973, 39, 367-373.

Watson, P. I.Q.: The racial gap. Psychology ... ...72, 6, 48-50, 97, 99.

Wechsler, D. Manual for the Wechsler intelli... ...Scale for children-revised. New York: The Psychological Corpor... ...1974.

Williams, R. Danger: Testing and dehumanizin... ...hildren. Clinical Child Psychology Newsletter, 1970, 9, 5-6.

BEST COPY AVAILABLE

101X

and conflicts of the child not primarily related to family malfunction or marital conflict. Working concurrently, or sequentially with family and individual therapy can be more effective than either one singularly.

An example of shifting from one modality to another might be that of an angry alienated adolescent whose problems initiated from family conflicts. Family sessions would deal with the nature of family conflicts, with the ways they affect family members, and with the resolution of conflicts through the learning and use of adaptive ways for the family to act and communicate. However, it does not follow that the adolescent's anger and dysfunctional behavior will necessarily cease. The resolution of family conflict we consider to be necessary for behavioral change, in some cases is not sufficient. Behaviors may be maintained by factors other than those which initiated them. A shift to individual or group sessions might best deal with residual problems, as well as with difficulties not associated with family conflict.

Another variation, the concurrent use of individual and family sessions, works well when specific aims are kept in mind for each modality. In cases where children have suffered insults to their self esteem as a result of physical illnesses, disabilities or learning impairments, they may do well in individual or group therapy which runs concurrent with family therapy. Family sessions would focus on the family's emotional reaction to the disability, as well as on those parental behaviors which impede their children's development. Individual or group sessions might focus on helping these children develop problem solving skills with their peers or in school.

Alternating modalities may be beneficial in cases of marital discord. In family sessions parent-child conflicts would receive attention, whereas in marital sessions sexual, financial and intimate relationship problems are aired.

SUMMARY

Our view of family therapy has evolved primarily from our clinical experience and not from research or theory. While our position fits closest to general social learning theory (Bandura 1977), we utilize diverse approaches in our work with children and families. We feel that too often the form of therapy used has been determined by past training and by therapist personality variables rather than by the needs of the particular patient. If psychotherapy is to mature as a science and an art, practitioners must realize that different problems require different treatments.

REFERENCES

Bandura, A. *Social learning theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1977.

Bateson, G., Jackson, D., Haley, J. & Weckland, J. Towards a Theory of Schizophrenia. *Behavioral Science*, 1956, *1*, 251-264.

Garfield, S., & Kurtz, R. A Survey of Clinical Psychologist: Characteristics, Activities, & Orientation. *The Clinical Psychologist*, 1974, *28*, 7-10.

Glenn, J. The adopted Theme in Edward Albee's "Tiny Alice" and "The American Dream." *The psychoanalytic study of the child*, 1974, *29*, 413-429.

Gurman, A. S. & Kniskern, D. P. Research on Marital and Family Therapy: Progress, Perspectives, and Prospects. In S. L. Garfield and A. E. Bergin (Eds.) *Handbook of psychotherapy and behavior change: an empirical analysis*. New York: John Wiley and Sons, 1978.

McDermott, J. J. and Char, W. F. The Undeclared War Between Child and Family Therapy. *Journal of the American Academy of Child Psychiatry*, 1974, *13*, 422-436.

Satir, V. *Conjoint family therapy*, Palo Alto: Science and Behavior Books, 1967.

Requests for reprints should be sent to Raymond DiGiuseppe, Ph.D., Long Island Jewish-Hillside Medical Center, 15-26 Hassock St., Rockaway, New York 11691.

Professional Child Abuse: How Children are Abused While Being Helped

Louise Mead Riscalla

Dr. Riscala is a Clinical Psychologist at Greystone Park Psychiatric Hospital. Her articles on rehabilitation, psychological assessment, psychosomatic medicine, attitudes of professionals toward "patients," and children's rights have appeared in numerous journals.

ABSTRACT

The professional as a child abuser continues to be a neglected area of child abuse because it frequently occurs in the process of helping children and in subtle ways that tend to be overlooked. This article examines some of the ways that professionals abuse children. It is essential that professionals become aware of their own motivations and consequences of their actions in order to prevent child abuse.

Child abuse arises from ambivalence manifested by loving, caring feelings as well as hateful feelings toward children. This ambivalence toward children occurs not only in parents, but permeates institutions, governments, and professions. The professional as a child abuser continues to be neglected because abuse often occurs in the process of helping children and in subtle ways that tend to be overlooked or are not considered as being abusive. Some of the ways that professionals abuse children will be explored with the intent that professionals will think more about their own motivations and what they are really doing while trying to help children.

Psychotrophic drugs may have a beneficial effect. However, it has been shown that drug-induced changes in behavior do not continue when the medication is discontinued (Shaffer, Costello, and Hill, 1968; Weiss et. al, 1975). Hyperactivity is frequently treated by psychotrophic drugs with the risk that these drugs could mask physical symptoms, thereby jeopardizing health. Walker (1974) reported a case of an 8-year-old boy with tantrums who had been treated by another physician with psychotrophic drugs who was found to be pre-diabetic. A 5-year-old girl with a history of disruptive behavior had seen a number of physicians until it was finally discovered that she had poor oxygenation due to a cardiac condition (Walker, 1974). Physicians are often unable to observe a child's behavior at home or at school with the result that drugs are prescribed on the basis of the complaints of parents and teachers. Parents and

Vol. 9, No. 1

2:3 BEST COPY AVAILABLE

teachers are inclined to become so fed up with a child's behavior that they seek immediate relief for themselves by having the child medicated.

It is well known that psychotherapy requires a combination of training, experience, and professional expertise. Administrators of institutions and community agencies under pressure to provide psychotherapy who cannot afford the cost of employing trained, experienced therapists, are often compelled to utilize personnel who are unqualified to do psychotherapy, with the risk of perpetuating existing pathology or creating additional problems where none previously existed. In order to avoid the possibility of legal involvement, including malpractice suits, psychotherapy is frequently practiced under different headings so that, in effect, anyone can do psychotherapy.

"Institutional and governmental abuse is all the more pernicious, because a cloak of official sanction serves to protect and perpetuate it (Grainger, Brant, & Brant, 1976, p. 171)." The state, through courts, acting in the best interests of children, frequently charges parents with child neglect or abuse. Children are then removed temporarily into a foster home, children's shelter, hospital, etc., where they are often abused and generally live in insecurity and fear of their disposition. Children often prefer to be in their own neglecting home situation because they have a considerable amount of loyalty and devotion to their parents in spite of being mistreated. "The child cannot plead the Fifth Amendment and what he unwittingly reveals about himself will be used when the court is trying to decide what to do with him. This seems to be a shocking invasion of privacy and one that again, is usually justified as being in the child's best interest (Tooley, 1970, p. 487)."

Children have the right to counsel, yet have complained that they cannot afford representation and are only seen by a public defender for 5 or 10 minutes before a hearing which can affect the outcome of their lives. If a child complains, there is little or no recourse because on paper, the child has had representation and the right to counsel has been implemented according to the law. Over-zealous child advocates anxious to make a name for themselves or their "cause" often latch on to issues and create controversy where the child is caught in the middle of a power struggle. Useful programs are forced to be abandoned thereby depriving children of valuable services because of the interpretation by well-meaning child advocates who are unable to provide the so-called better alternative programs they recommend.

There is a tendency to erroneously believe that to solve or prevent the problem of child abuse, something must be done about it such as to provide additional trained staff, more funds, change the system by doing something to improve it, etc. Services, money, and trained staff alone do not cure or prevent child abuse. The abuse of children by professionals and the prevention of this abuse depends on human qualities, and what is done for children on paper or in action is a by-product of these qualities. Professionals are generally reluctant to admit that they abuse children. However, an awareness of the ways that children are abused in the process of being helped is basic toward the prevention of child abuse by professionals. Honest, self-examination, including an open-minded understanding of what is going on in the context of day-to-day experiences, an awareness of one's motives, and a perception of the consequences of actions taken are essential. The unconditional expression of qualities such as intelligence, patience, honesty, an affirmation of the child's right to be what the child is, etc., do not require any expenditures of money and have a beneficial impact on children so that children are actually helped rather than abused.

REFERENCES

Grainger, G., Brant, R. and Brant J. Protecting children and families from abuse. In G. Koocher (Ed.) *Children's rights and the mental health professionals*. New York, John Wiley and Sons, 1976.

Shaffer, D., Costello, A., and Hill, I. Control of enuresis with imipramine. *Archives of Diseases in Children*, 1968, *43*, 665-671.

Tooley, K. Ethical considerations in "involuntary commitment" of children in psychological testing as part of legal procedures. *Mental Hygiene*, 1970, *54*, 484-489.

Walker, S. III. Drugging the American child. We're too cavalier about hyperactivity. *Psychology Today*, 1974, *8*, 43-48.

Weiss, G. et. al. The effects of long-term treatment of hyperactive children with methylphemidate. *Journal of the Canadian Medical Association*, 1975, *112*, 159-165.

Requests for reprints should be sent to Louise Mead Riscalla, 8 Lahiere Avenue, Edison, New Jersey 08817.

DEATH and CHILDREN

THE FACTS OF DEATH ARE AS IMPORTANT TO CHILDREN AS THE FACTS OF LIFE! This issue is an awareness-enhancing examination of the myriad facets of the still taboo topic of death and children.

Topics include: children's concepts of death; conversations with children about death; death education for children and youth; pre-adolescent suicide; protecting children from life-threatening parents; bereaved children; fatally ill children; emotional support programs for parents of dying children.

Articles by Feifel, Leviton, Kastenbaum, Kubler-Ross, Paulson, Lifton and other distinguished child workers and thanatologists.

See order blanks on pages 89 to 91.